HP Prime
AP Statistics
Summer
Institute

Materials by GT Springer
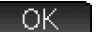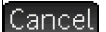
Table of Contents

Getting Acquainted With HP Prime

HP Prime is a color, touchscreen graphing calculator, with multi-touch capability, a Computer Algebra System (CAS), an Advanced Graphing app that lets you graph any relation in two variables (graphing something like $\sin(xy) = \cos(xy)$ for example), and a set of three apps for statistics (Statistics 1Var, Statistics 2Var, and Inference). In this section, we'll take a look at how to find your way around HP Prime, and get acquainted with the Prime app structure.

First, here are a few conventions we'll use in this document:

- A key that initiates an un-shifted function is represented by an image of that key: ⌨, ⌨, ⌨ and so on.

- A key combination that initiates a shifted function (or inserts a character) is represented by the appropriate shift key (⌨ or ⌨) followed by the key for that function or character: ⌨ ⌨ initiates the natural exponential function and ⌨ ⌨ inserts the letter F.

- The name of the shifted function may also be given in parentheses after the key combination: ⌨ ⌨ (Clear), ⌨ ⌨ (Plot Setup)

- A key pressed to insert a digit is represented by that digit: 5, 7, 8, and so on.

- All fixed on-screen text—such as screen and field names—appear in bold: **CAS Settings**, **Xstep**, **Decimal Mark**, and so on.

- A menu item selected by touching the screen is represented by an image of that item: Sto ▸ , OK , Cancel , and so on.

  NOTE: You must use your finger to select a menu item, or navigate to the selection and press ⌨ .
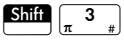
- Cursor keys are represented by ⌄, ⌃, ▷ , and ⌃. You use these keys to move from field to field on a screen, or from one option to another in a list of options.

The ON-OFF key is at the bottom left of the keyboard. When a new HP Prime is turned on for the first time, a "splash" screen appears that invites the user to select a language and to make some initial setup choices. For most users, accepting the default options is the way to go.

The screen brightness can be increased by pressing and holding ⌨ and ⌨ or decreased by pressing and holding ⌨ and ⌨ .

Take a minute to look at the layout of the keyboard. The top group of keys, with the black background, is primarily for navigating from one environment to another. Pressing ⌨ takes you to the home calculation screen, and pressing ⌨ takes you to a similar calculation environment for doing symbolic or exact computations. Pressing ⌨ takes you to a menu where you can select from all the applications in the HP Prime, like **Statistics 1Var** or **Inference** or **Function**. The bottom group of keys is mainly for entering or editing mathematical expressions. There are also environments for entering lists, ⌨ ⌨ , matrices, ⌨ ⌨ , and user programs, ⌨ ⌨ .

Some care was taken when deciding where to place certain keys. The number π for instance, is [Shift] [3]. The list delimiters, {}, appear just to the right of the LIST key, [Shift] [8] and the matrix delimiters, [], appear just to the right of the MATRIX key, [Shift] [5].
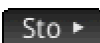
Things you can do in both CAS and Home views:
- Tap an item to select it or tap twice to copy it to the command line editor
- Tap and drag up or down to scroll through the history of calculations
- Press [Menu Paste] to retrieve a previous entry or result from the other view
- Press the Toolbox key ([Mem B]) to see the Math and CAS menus as well as the Catalog
- Press [Units c] to open a menu of easy-to-use templates
- Press [Esc Clear] to exit these menus without making a selection
- Tap [Sto ▶], [Copy], and [Show] menu buttons

Home View

Turn on your HP Prime and take a look at the different sections of the screen in the **HOME** view. The top banner across the top is called the Title Bar, and it tells you what operating environment you are currently working in (like **HOME** or **FUNCTION SYMBOLIC VIEW**). If you press a shift key, an annunciator comes on at the left of the Title Bar. On the right, you see a battery level indicator, a clock, and the current angle mode. You can tap this Quick Settings section at the top right to see a calendar (by tapping the date and time), connect to a wireless classroom network (by tapping the wireless icon), or change the angle mode (by tapping the angle mode indicator).
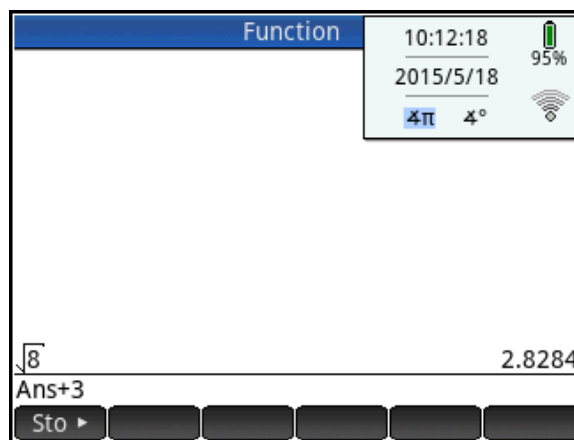
The middle section of the **HOME** view contains a history of past calculations. You can navigate through the history with the cursor keys, or using your finger to select (by tapping) or scroll (by swiping). The edit line is just below the history section. This is where you enter mathematical expressions to evaluate numerically. At the bottom are the menu keys, consisting of [Sto ▶] in the **HOME** view. These menu keys are context sensitive. Their labels and function changes depending on the environment you're in.

Home view is for numerical calculations. Press [Settings] to open Home view if you are not already there. Press [Mem B] to open the Toolbox menu and tap [Math]. From the math menu, tap Probability, Cumulative, and select Normal. The NORMAL_CDF() function will be pasted into the command line. Between the parentheses, enter 0, 1, -1, 1, as shown in the figure to the right. Press [Enter ≈] to see the expression evaluated numerically. The figure shows a few other calculations as well.
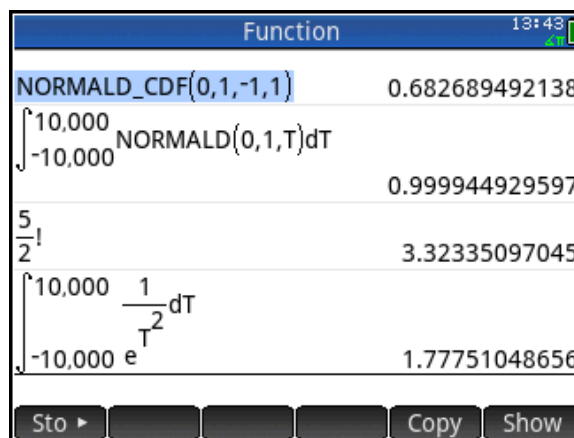
In all of these cases, the results evaluate to a real number. In Home view, all results evaluate to a real or a complex number, or a matrix, list, etc. of real and/or complex numbers. You can tap on any previous input or result in the history to select it. When you do, two new menu buttons appear: Copy and Show. The former copies the selection to the command line at the cursor position while the later typesets the selection in textbook format in full-screen mode.

CAS View

CAS view, on the other hand, is for symbolic or exact numerical results. Press CAS Settings to open CAS view. Let's repeat our first integral calculation. Press Units and choose the integral template. For the lower limit, enter -∞. Press

Shift 9 to find the infinity symbol. Enter ∞ for the upper limit. For the integrand, press c and tap Math. From the Math menu, tap Probability, Density, and select Normal. Complete the command as shown
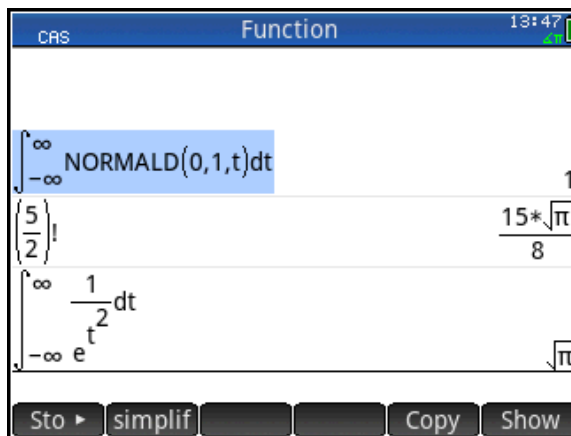
and press Enter. As shown on the figure, the result is an exact numerical value. Similarly, $\frac{5}{2}!$

and $\int_{-\infty}^{\infty} \frac{1}{e^{t^2}} dt$ evaluate exactly.

Copy and Paste

As mentioned earlier, both the CAS and Home view histories use Copy to copy the selection to the command line. There is also Shift View/Copy (Copy) and Shift Menu/Paste (Paste) that copies the selection to the Prime clipboard and pastes from that clipboard to the cursor position. This functionality makes it possible to copy and paste from one environment to another anywhere in your HP Prime. With data, you can tap and hold, then drag to select a rectangular array of cells, then copy and paste anywhere else. With the HP Prime Virtual Calculator, you can copy an array of cells in a spreadsheet on your PC and paste the numerical data anywhere in the Prime Virtual Calculator. You can then send the data to another HP Prime.

Delete and Clear

In both CAS and Home views, you can select an item in history and press Del to delete it. Press Shift Esc/Clear (Clear) to delete the entire history. If you are in any view (Symbolic, Plot Setup, etc.), Shift Esc/Clear (Clear) will reset (clear) all settings in the current page of the view back to their factory defaults.

HP Prime Apps and Their Views

The HP Prime graphing calculator comes pre-loaded with a number of apps. Each app was designed to explore and area of mathematics or to solve problems of a specific type. Every Prime app is divided into one or more views. Most commonly, an app has a Symbolic view, a Plot (or Graphic) view, and a Numeric view. In this sense, the apps all have a common structure so they are easier to learn to use as a set of apps. The Prime app schema is shown in the figure below.

HP Apps and their Views

Symbolic

Graphic(Plot)

Numeric
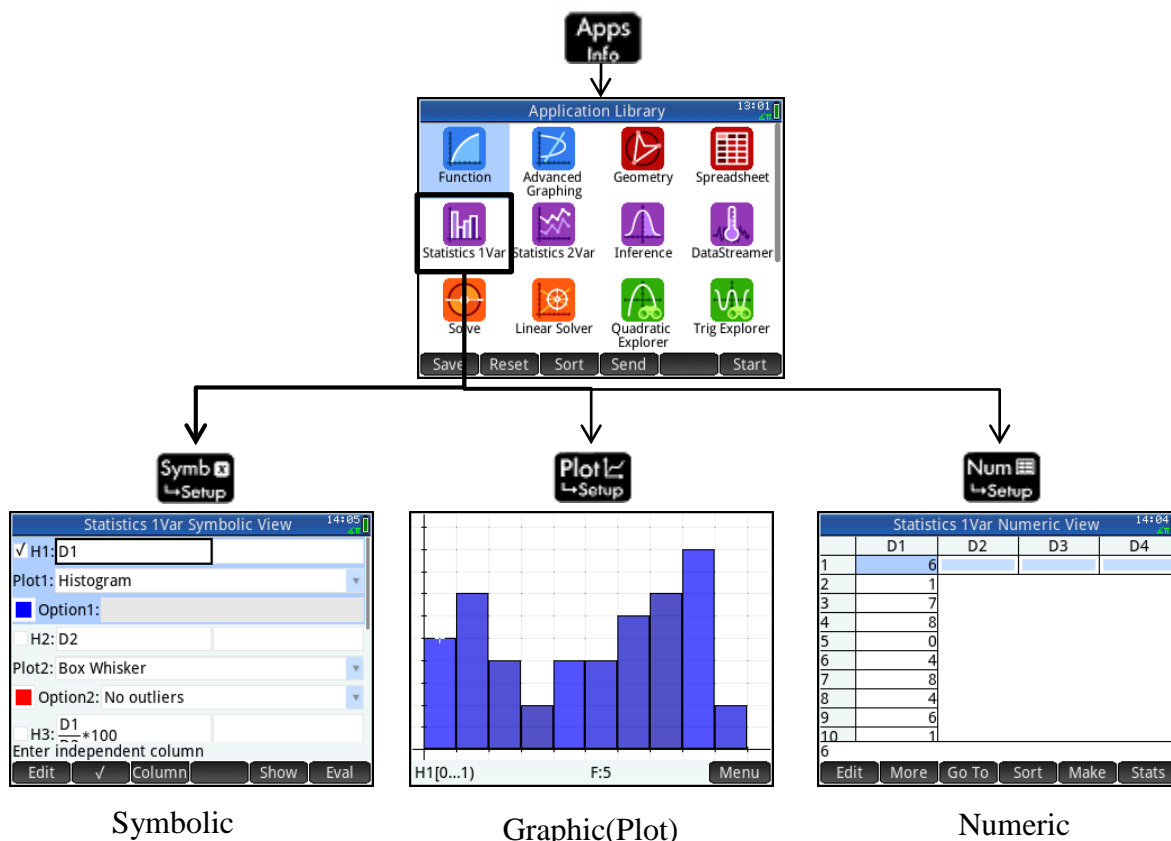
Press ![Apps Info] to open the App Library. Tap on an app to start it, or navigate the library using the cursor pad and tap ![Start] to launch the app.

Fill the app with data while you work; you can come back to your saved app anytime-even send it to your colleagues! You can save an app with a name you'll remember; then reset the original app and use it for something else. HP Apps have app functions as well as app variables; you can use them while in the app, or from the CAS view, Home view, or in programs.

Women in Computer Science

There is a concern that women are under-represented in the high-tech field of computer programming. In this activity, we explore the percentage of bachelor and advanced degrees in Computer Science awarded to women in various countries.

HP Prime Functionality Introduced:

Using the Statistics 1Var app Numeric, Symbolic, and Plot views; calculating summary statistics for 1-variable data sets;

AP Statistics Content:

Constructing and interpreting graphical displays of univariate data (stemplot, histogram); summarizing distributions of univariate data; using boxplots

The table to the right shows the percentage of Computer Science degrees awarded to women in 22 countries in the year 2011.

1. Press **Apps Info** and select `Statistics 1Var` and enter the data manually in list D1 in Numeric view. Or get the *WomenInCS* app from your instructor.

2. Tap **Sort** and then tap **OK** to sort the data in ascending order. The pinch vertically to decrease the font size. The data set has 22 values; by inspection, what is the 5-number summary of the data set?

Minimum=_____

Q1=_____

Median=_____

Q3=_____

Maximum=_____

2011 CS degrees awarded to women			
Country	Percent	Country	Percent
Austria	15.7	Israel	26.1
Belgium	9.8	Italy	25.5
Czech Rep.	12.2	Netherlands	12.8
Denmark	26.8	Norway	13.2
Estonia	23.1	Poland	15.6
Finland	23.9	Portugal	22.4
France	16.6	Spain	17.0
Germany	16.7	Sweden	29.4
Hungary	17.2	Switzerland	8.6
Iceland	13.1	UK	19.0
Ireland	42.3	USA	21.1

3. Press ⌨ to open Symbolic view. Set H1 to use D1 as its data and to draw a stem and leaf plot, as shown in the figure to the right. You can type in D1 or tap Column and select D1 from the list. Tap on the **Plot1** field to select it and then tap again to open the list of plot options and select Stem and Leaf.

4. Press ⌨ to return to Numeric view and tap Stats to calculate summary statistics for H1.

5. Does the 5-number summary agree with the values you gave in #2?
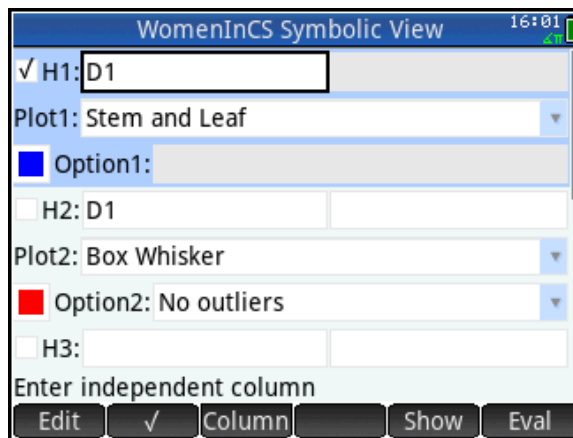
6. The mean is a bit larger than the median. What does this mean about any potential skew of the data?

7. Press ⌨ and select Autoscale to see the stem and leaf plot. Tap on a data point or use the cursor pad to navigate the data set.

8. From the 5-number summary and the stem plot, what do you conclude about the participation of women in Computer Science degree programs?
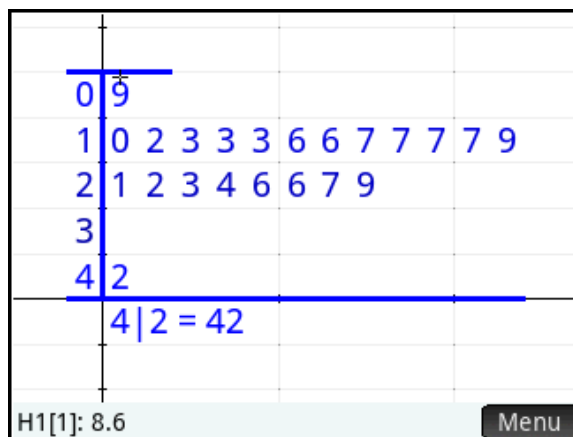
9. Return to Symbolic view and change the plot to a histogram.

10. Press ⇧Shift 🔲Plot to open Plot Setup view. Change the settings to agree with the figure shown to the right.

WomenInCS Plot Setup		14:25
H Width: 5		
H Rng: -5	60	
X Rng: -5	60	
Y Rng: -1	8	
X Tick: 5		
Y Tick: 1		

Enter vertical tick spacing

| Edit | | Page ½ ▼ | | |
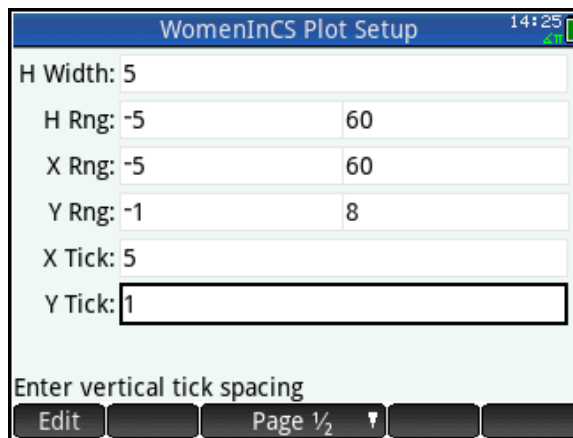
11. Press 🔲Plot to see the histogram. How would you describe this distribution?

_____

_____

_____

H1[15...20)	F:7	Menu
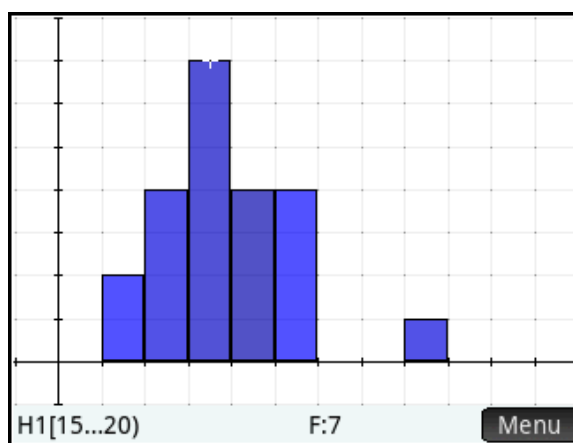
12. Is the maximum an outlier? Return to Symbolic view and change the **Plot1** field to Box Whisker. In the **Option** field, select Show outliers.

13. Press 🔲View and select Autoscale to see the plot.

14. The data point of 42.3 from Ireland is shown as an outlier. What can we learn about the participation of women in the field of computer programming from this outlier?

H1[Outlier]: 42.3	Menu
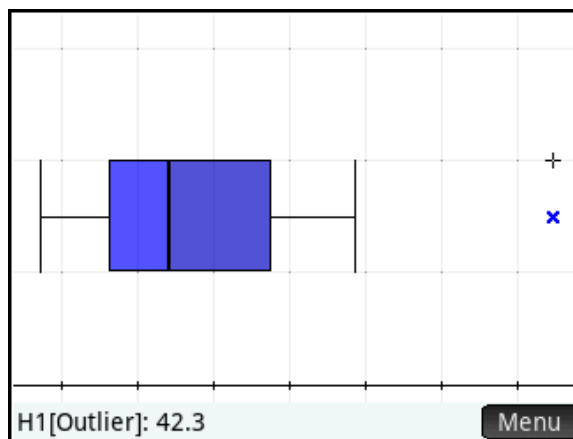
Answers

2. The minimum is the first data point, 8.6; Q1 is the 6th data point, 13.2; the median=17.1, the average of the 11th and 12th data points; Q3 is the 17th data point, 23.9; the maximum is the last data point, 42.3

5. Yes, the computed 5-number summary matches our answers from #2

6. The mean being greater than the median indicates a slight positive skew.

8. Since over half the data points are less than 20%, it seems clear that, at least in Europe, the UK, and the USA, women are indeed under-represented.

11. The distribution is unimodal and fairly symmetric, with a possible slight positive skew.

14. The fact that Ireland awarded 42.3% of CS degrees to women indicates that it is possible to have gender equality in this field. India also reported more than a 40% participation rate in CS degree programs in 2011.

The data in this activity come from the *2014 Digest of Education Statistics*, developed by the National Center for Educational Statistics. The source of the data is the Organization for Economic Cooperation and Development's *Online Education database*. You can find the data here:
https://nces.ed.gov/programs/digest/d14/tables/dt14_603.70.asp

Anscombe's Quartet

Anscombe's Quartet refers to four data sets devised by Francis Anscombe in 1973. In this activity, the data is already loaded into an app called *Anscombe*.

HP Prime functionality introduced:

Using the Statistics 2Var app Numeric, Symbolic, and Plot views; calculating summary statistics for 2-variable data sets; the Resid() command

AP Statistics Content:

Analyzing patterns in scatter plots; correlation and linearity; least-squares regression lines; residual plots, outliers, and influential points

Part 1

1. Press **Apps Info** and scroll down to the *AnscombeQuartet* app. Tap **Start** to start the app.

2. Press **Num Setup** to see the data sets in Numeric view. C1 holds the common x-variable data for the 1st three datasets. C2, C3, and C4 each contain a different y-variable data set for the 1st 3 data sets. The 4th dataset is contained in C5 and C6.

3. Tap **Stats**. The summary statistics are displayed, as shown to the right.

4. Tap **X** to see summary statistics for the independent variable.

5. Tap **Y** to see similar statistics for the dependent variable. Tap **OK** when you are done.

	C1	C2	C3	C4	C5	C6	C7
1	10	8.04	9.14	7.46	8	6.58	
2	8	6.95	8.14	6.77	8	5.76	
3	13	7.58	8.74	12.74	8	7.71	
4	9	8.81	8.77	7.11	8	8.84	
5	11	8.33	9.26	7.81	8	8.47	
6	14	9.96	8.1	8.84	8	7.04	
7	6	7.24	6.13	6.08	8	5.25	
8	4	4.26	3.1	5.39	19	12.5	
9	12	10.84	9.13	8.15	8	5.56	
10	7	4.82	7.26	6.42	8	7.91	
11	5	5.68	4.74	5.73	8	6.89	
12							

Enter value or expression

Edit More Go To Sort Make Stats

Anscombe Numeric View

	S1	S2	S3	S4
n	11	11	11	11
r	0.8164	0.8162	0.8163	0.8165
R^2	0.6665	0.6662	0.6663	0.6667
sCOV	5.501	5.5	5.497	5.499
σCOV	5.0009	5	4.9973	4.9991
ΣXY	797.6	797.59	797.47	797.58

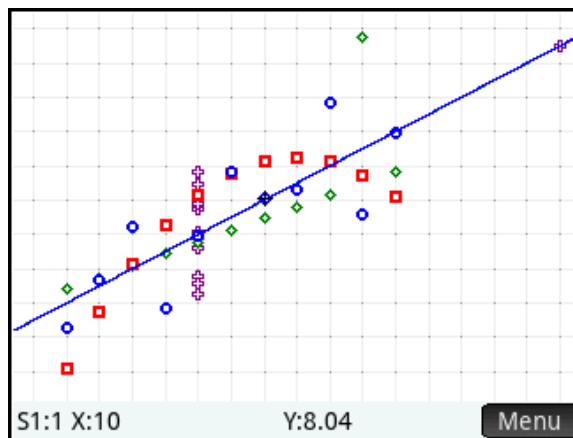Number of items

More Stats• X Y OK

6. Write a sentence or two about what you have discovered so far about these four data sets.

7. Press ⬜ and select `Autoscale`.
8. Each scatter plot is color-coded. How many scatter plots can you distinguish?

_____

9. How many linear fits can you distinguish?

_____

10. Press ⬜ to open Symbolic view.

As you can see, S1 has been set to use C1 as the independent variable and C2 as the dependent variable. Likewise, S2 uses C1 for the independent variable and C3 for the dependent variable. Finally, S3 uses the common C1 for the independent variable and C4 for the dependent variable. Not shown is S4, which uses C5 and C6. By default, all 4 of these analyses have been checked to be active.
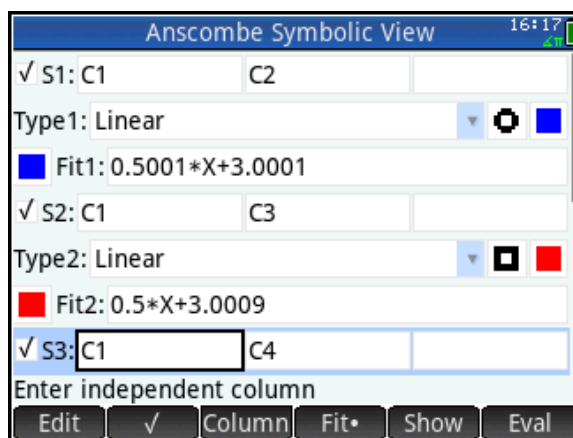
11. Examine the fits for each of the analyses S1–S4. Why did you only see one fit in Plot view?

_____

12. Uncheck S2, S3, and S4, leaving only S1 checked.
13. Press ⬜ to open Plot view. The scatter plot and linear fit of S1 are displayed.
14. Press ⊙ to move the tracer from the scatter plot to the fit line. Tap `Menu` and `Defn` to see the linear fit expression in X: S1=0.5*X+3.
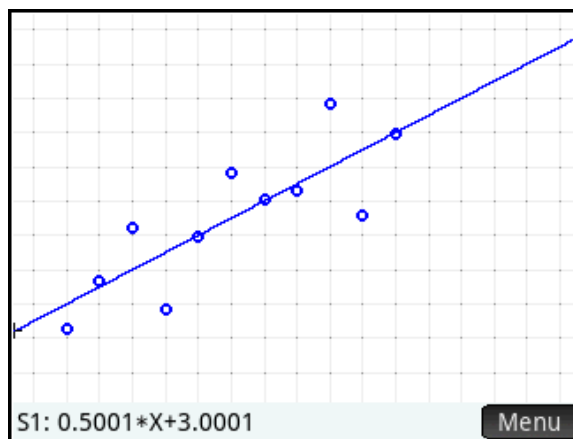
Let's look at a residual plot for S1. To create the residual plot, we will store the residuals in list C7 and then create a scatter plot with C1 and C7.

15. Press ⌂ (Settings) to open Home view and enter Resid(S1)▸C7. To find the Resid command, press 📖 (Mem B), tap App, then Anscombe and select Resid.

16. Press (Symb Setup) to open Symbolic view. Change S2 to use C1 and C7, as shown to the right.

17. Press (View Copy) and select Autoscale. Do you see any pattern to the residual plot?
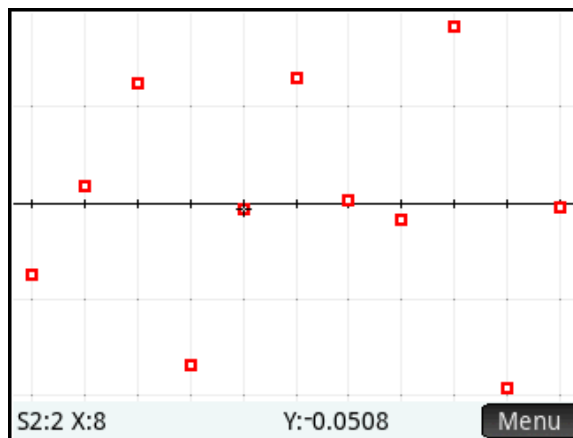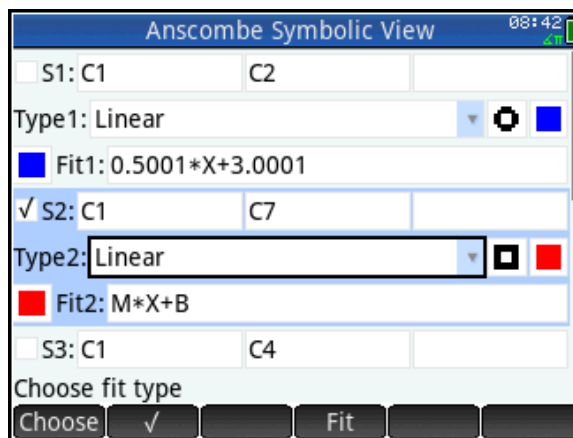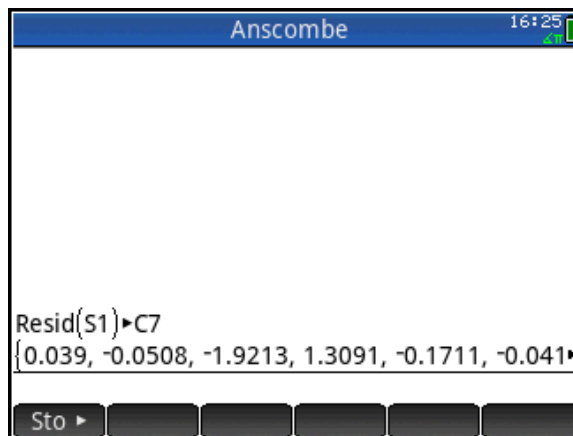
18. For the first data set, does a linear model seem appropriate?

_____

_____

Part 2
Repeat the previous procedures with the
second data set. That is, make S1 use C1
and C3 to create a scatter plot with those
lists. Then store the residuals for S1 in C7
and create a residual plot in S2, using lists
C1 and C7.

19. Do you detect any pattern in the
    scatter plot of the data or the
    residuals?

S1:1 X:10   Y:9.14   Menu

S2:4 X:9   Y:1.2691   Menu

_____

_____

_____

_____
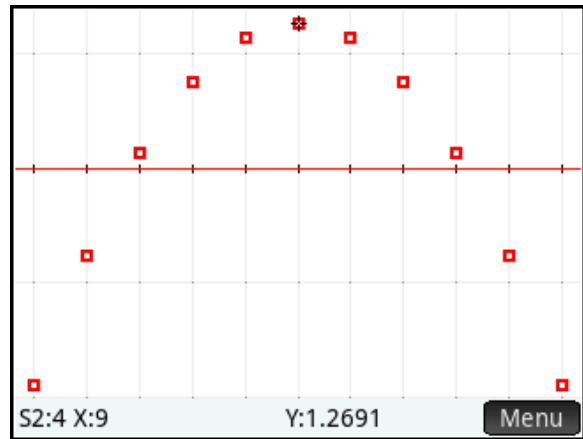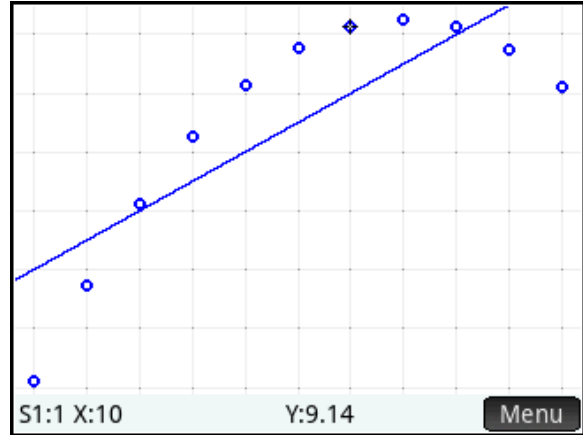
20. Is a linear model appropriate for the
    second data set?

_____

21. Examine the third data set the same way you examined the first and second data sets.
    Record your conclusions in the space below.

_____

_____

22. Examine the fourth and record your conclusions in the space below.

_____

_____

Answers

6. The four data sets have remarkably similar summary statistics. Note especially that the correlation coefficients all round to 0.82.

8. Four scatter plots can be easily detected: one blue, one red, one green, and one purple.

9. Only one fit, a blue line, can be seen.

11. Only one fit could be seen because the four fit lines are almost identical.

17. There is no detectable pattern for the residual plot of the first data set.

18. Yes, a linear model seems appropriate for the first data set.

19. Yes, there is a clear quadratic pattern to both the scatter plot and the residual plot for the second data set.

20. No, a linear model is not appropriate here.

21. The third data set is clearly linear, though a different linear than the fit line. The influence of an outlier is illustrated in this data set.
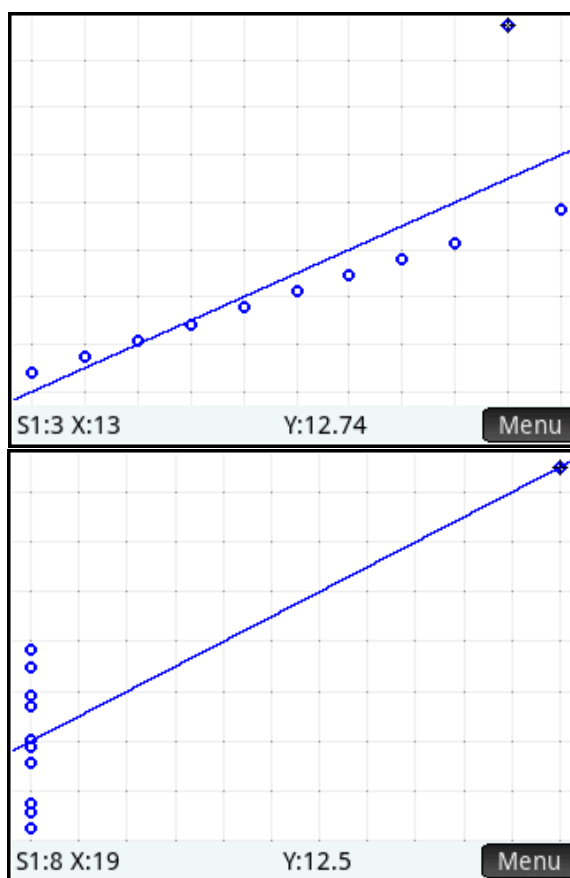
22. The fourth data set is not linear at all. This data set shows that an outlier can generate a high correlation value even though the relationship is simply not linear.

Anscombe's Quartet was devised to stress the importance of looking at data graphically as the first step in any analysis!

Cigarette Smoking in the United States

In this activity, we look at data from the Center for Disease Control on the number of daily smokers in the United States and the trend in that data. You can start with either the *Statistics 2Var* app or get the *Smokers* app from your instructor.

HP Prime Functionality Introduced:

Using the Statistics 2Var app Numeric, Symbolic, and Plot views; predicting y-values from x-values using a fit; using the Inference app's inference for regression methods

AP Statistics Content:

Least-squares regression lines; confidence interval for the slope of a least-squares regression line; test of significance for the slope of a least-squares regression line

Part 1

The table to the right shows the percentage of people in the United States who smoke cigarettes daily, for each of the years from 2000 through 2010.

1. Press ⌨ to open the App Library and select the *Statistics 2Var* or *Smokers* app.

2. If you use the Statistics 2Var app, enter the data into C1 and C2 of Numeric view; if you got the Smokers app from your instructor, then the data is already entered for you. If you enter the data manually, consider using 0 for 2000, 1 for 2001, and so on, as was done in the figure to the right.

3. What is the average rate of decline in the percentage of daily smokers annually from 2000 through 2010? Is a linear model appropriate?

Percentage of daily smokers in the USA	
Year	% of daily smokers
2000	17.7
2001	17.4
2002	17.8
2003	16.9
2004	15.8
2005	15.3
2006	14.9
2007	14.5
2008	13.4
2009	12.8
2010	12.4

4. Press [Symb/Setup] to open Symbolic view. Set S1 to use C1 as the independent data and C2 as the dependent data, with a linear fit. You can also choose colors for the scatter plot and fit. In the figure to the right, the scatter plot will be drawn in red and the fit will be blue.
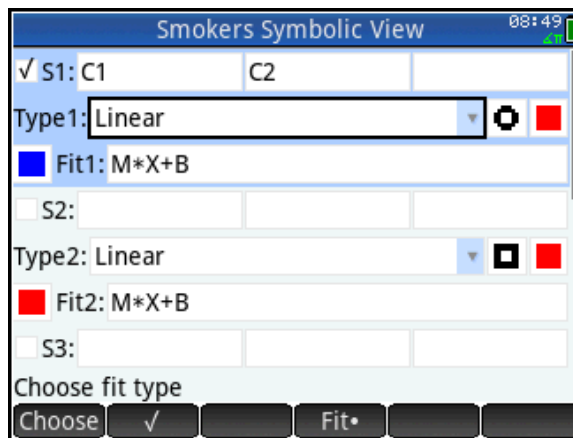
5. Press [View/Copy] and select Autoscale. Use your fingers to pinch and drag until you can see both axes and the x-intercept of the fit.

6. Tap [Menu] and [Defn] to see the linear fit expression in X: S1= -0.58*X+18.2545.

7. Explain what the parameters of this fit equation mean in terms of the percentage of daily smokers in the USA.

8. The PredX() command uses our fit to calculate the x-value for a given y-value. As you can see in the figure to the right, PredX(0)=31.4734. Explain what this means in terms of the percentage of daily smokers in the USA.

_____

_____
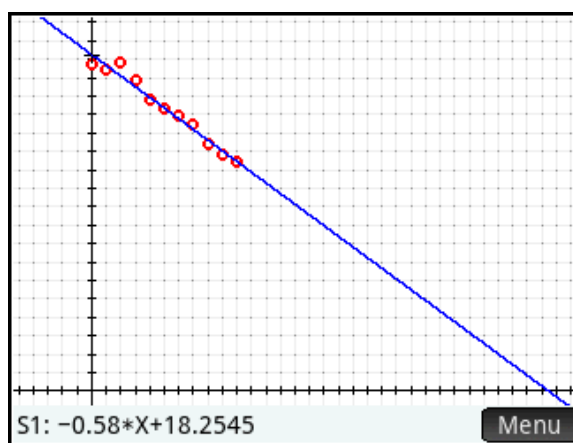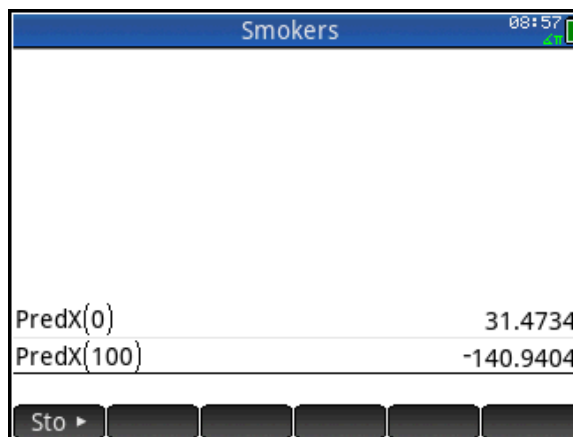
9. PredX(100)= -140.9404. Explain what this means and also express your confidence in this result.

Part 2

In this part of the activity, we perform a linear t-test and construct a 95% confidence interval for the slope of the line. First, let's suppose we want to perform a t-test on the slope at the $\alpha=0.05$ level.

1. Copy the data in Numeric view. To do this, tap and hold on the first data point in C1; then drag below and to the right to the last data point in C2. With all the data selected, press **Shift** **View/Copy** (Copy).

2. We will now perform the linear t-test. Go to the App Library and select the Inference app. The app opens in Symbolic view. Tap the **Method** field and select `Regression`. In the **Type** field, select `Linear t test`.

3. The null hypothesis for the linear t-test is that the slope of the regression line is zero. Since the data indicate that the percentage of smokers is decreasing, our alternative hypothesis is that the slope of the regression line is negative. In Symbolic view, select $\beta_1 < 0$ in the **Alt Hypoth** field.

4. Press **Num/Setup** to open Numeric view and paste your smokers data by pressing **Shift** **Menu/Paste** (Paste). You will be prompted to paste your data as either grid data or text. Choose grid data and tap ⬜OK⬜.

5. Tap ⬜Calc⬜ to view the results of the linear t test. Tap OK when you are done viewing the test results.

6. What does the p-value indicate about the null hypothesis vs. the alternative hypothesis?

Inference Symbolic View	09:06
Method: Regression	▼
Type: Linear t test	▼
Alt Hypoth: $\beta_1 < 0$	▼

Choose the alternative hypothesis

⬜ ⬜Choose⬜ ⬜ ⬜ ⬜ ⬜

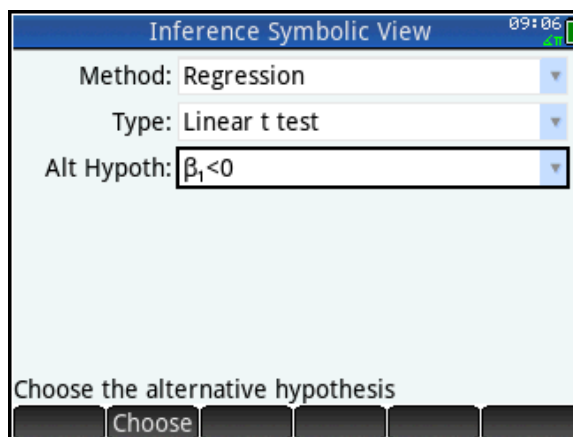| Inference Numeric View		09:09
	Xlist	Ylist
1	0	17.7
2	1	17.4
3	2	17.8
4	3	16.9
5	4	15.8
6	5	15.3
7	6	14.9
8	7	14.5
9	8	13.4
10	9	12.8
0		

⬜Edit⬜ ⬜More⬜ ⬜Go To⬜ ⬜Import⬜ ⬜Make⬜ ⬜Calc⬜

| Stats		09:23
Test T	-16.2367	
P	2.8295ᴇ-8	
DF	9	
β_0	18.2545	
β_1	-0.58	
serrLine	0.3747	
serrSlope	0.0357	
serrInter	0.2113	
r	-0.9834	
R²	0.967	

Calculated test probability

⬜ ⬜More⬜ ⬜ ⬜ ⬜ ⬜OK⬜

7. Tap [OK] to return to Numeric
    view. Tap [Plot Setup] to open Plot view.
    You will see the scatterplot of the
    data, along with the linear fit.

8. Press D to move from the fit line to
    the scatterplot.

I1:1 X:0   Y:17.7   Menu
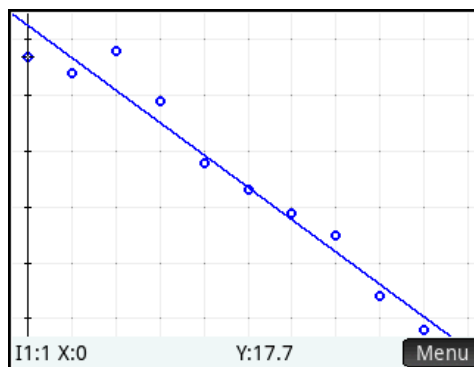
9. Press ⊙ again to see the
    scatterplot of the residuals. The
    scatterplot shows no discernable
    pattern. Press ⊙ to return to the
    scatterplot of the data and the fit.

I2:1 X:0   Y:-0.5545454545   Menu
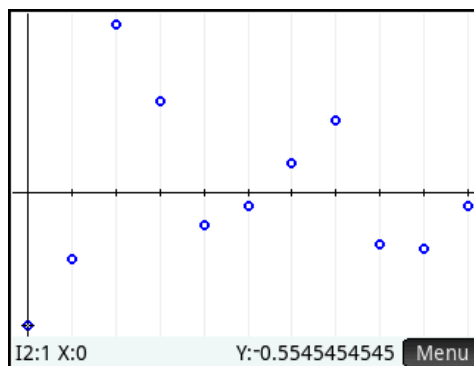
10. Press ⊙ once more to see the
    histogram of the residuals. You will
    have to set H Width manually to get
    meaningful results. Again, press ⊙
    if you wish to return to the
    scatterplot of the residuals.

I3[-0.6...-0.5)   F:1   Menu
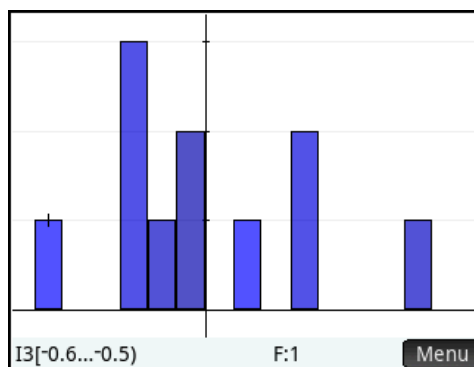
11. Finally, press ⊙ once more to see
    the normal probability plot of the
    residuals, which shows no alarming
    departure from linearity. Press ⊙
    once more to return to the original
    scatterplot of the data and the fit
    line, or press ⊙ to return to the
    histogram of the residuals.

H4: X:-0.5545454545 Y:-1.6906216296 Menu
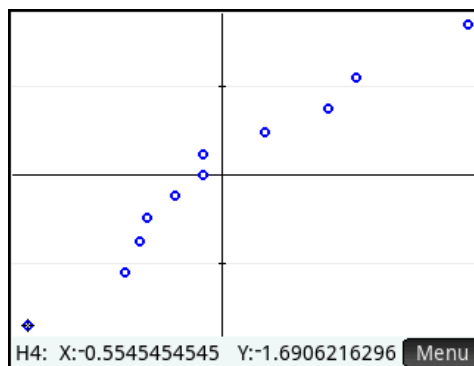
12. To construct a 95% confidence interval for the slope, return to Symbolic view and change the **Type** field to `Interval: Slope`.

13. Return to Numeric view and tap `Calc`. You will be prompted to enter a confidence level. Enter C=0.95 and tap `OK`.

14. The Results page for the confidence interval will be displayed, as shown in the figure to the right. What is the 95% confidence interval for the slope of the true regression line and what does it mean about the change in the percentage of daily smokers in the USA per year?

Results	12:44
C	0.95
Crit. T	2.2622
DF	9
β_1	-0.58
serrSlope	0.0357
Lower	-0.6608
Upper	-0.4992

0.95

	More				OK

Answers

Part 1

3. The average annual rate of change in the percentage of smokers in the USA is -0.53, indicating that there is a 0.53% decrease in the number of daily smokers in the USA each year.

7. The slope (m= -0.58) indicates an annual decrease in the percentage of daily smokers in the USA of 0.58%. The intercept (b=18.2545) indicates that in the year 2000, 18.2545% of the population of the USA were daily smokers. The latter is about half a percentage point higher than the data indicate.

8. PredX(0)=31.4734 can be interpreted to mean that there will be no more daily smokers in the USA sometime in the year 2031. This is an extrapolation.

9. PredX(100)= -140.9404 can be interpreted to mean that back in 1859, everybody in the USA smoked. This is an extreme extrapolation and clearly not a realistic conclusion. This can be seen by observing that Y>100 for X<-141. These values make no sense in the current context.

Part 2

6. Since p≈0.00000003, which is less than 0.05 (our α-level), we confidently reject the null hypothesis in favor of the alternative hypothesis that the slope of the true regression line is negative.

9. The 95% confidence interval for the slope of the true regression line is (-0.6608, -0.4992). We are 95% confident that the true decrease in the percentage of daily smokers in the USA is between 0.4992% and 0.6608%.

Teacher Notes

In general the Stats and Results pages from the Inference app (what you see when you tap `Calc` in Numeric view) include the statistics needed to calculate the main results by hand. For instance, in the linear t-test, the t-value we seek is given by $t = \dfrac{\beta_1 - \beta_{null}}{SE_{\beta_1}}$. The results page for the linear t-test, as shown in the figure, includes these values (β_{null} is known to be 0). Depending on the circumstances, one can go step-by-step to find the test t and its probability to verify the results shown.

Stats		09:23
Test T	-16.2367	
P	2.8295E-8	
DF	9	
β_0	18.2545	
β_1	-0.58	
serrLine	0.3747	
serrSlope	0.0357	
serrInter	0.2113	
r	-0.9834	
R²	0.967	

Calculated test probability

More   OK

Also note the standard error of the line (serrLine). This is the standard deviation of the residuals, $s = \dfrac{\sum (y_i - \hat{y}_i)^2}{n-2}$. It can be interpreted as saying that we will be off by 0.3747 on average when we predict the percentage of daily smokers from the year.

For the confidence interval of the slope of the regression line, the input confidence level is repeated for convenience and the critical value of t is given, followed by the degrees of freedom, the slope, and the standard error of the slope. These values are sufficient to calculate the confidence interval by hand, using the formula $\beta_1 \pm t^* \cdot SE_{\beta_1}$.

Results		12:44
C	0.95	
Crit. T	2.2622	
DF	9	
β_1	-0.58	
serrSlope	0.0357	
Lower	-0.6608	
Upper	-0.4992	

0.95

More   OK

The data in this activity come from the Center for Disease Control and Prevention, specifically, their *BRFSS Prevalence and Trends Data: Tobacco Use – Four Level Smoking Data for 1995-2000*. The data can be found here:
https://data.cdc.gov/Smoking-Tobacco-Use/BRFSS-Prevalence-and-Trends-Data-Tobacco-Use-Four-/8zak-ewtm

The Ebola Epidemic in West Africa

HP Prime Functionality Introduced:

Using the Statistics 2Var app Numeric, Symbolic, and Plot views; predicting y-values from x-values using a fit;
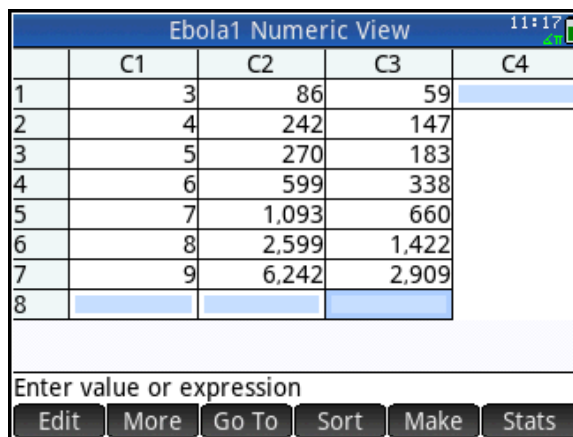
AP Statistics Content:

Analyzing patterns in scatter plots, least-squares regression lines, transformations to achieve linearity

Part 1

In March, 2014, the World Health Organization (WHO) began collecting data on the number of confirmed cases and deaths from Ebola in West Africa. Below are data from each month for the first six months. The data were collected on or about the 24th of each month. In this activity, we examine the data and build models based on the data.

Ebola Cases and Deaths in 2014		
Month	Cases	Deaths
3	86	59
4	242	147
5	270	183
6	599	338
7	1093	660
8	2599	1422
9	6242	2909

1. Either enter the data into Numeric view of the *Statistics 2Var* app or get the *Ebola1* app from your instructor.
2. To enter the data manually, press **Apps Info** to open the App Library and select `Statistics 2Var`. The app opens in Numeric view. Enter the data for the months in C1, the number of cases for each month in C2, and the number of deaths for each case in C3, as shown to the right.
3. By examining the data, explain why a linear fit is not appropriate.

If you are using the *Ebola1* app, it already has these data entered in Numeric view of the Statistics 2Var app. Just press **Apps Info** to open the App Library and select `Ebola1`.

4. Press ⌨ to open Symbolic view. By default, S1 should already be set to use C1 as the independent data and C2 as the dependent data. In **Type1**, select Exponential. Your Symbolic view should appear as in the figure to the right. You may also choose colors for your scatter plot and fit. Just tap twice on the color-picker for each one to open it. In the figure, we chose red for the scatter plot and blue for the fit.

5. Press ⌨ to return to Numeric view and tap Stats. Summary statistics for the S1 analysis are displayed. You can tap X and Y to see summary statistics for each of those variables alone. Note the correlation coefficient and its square, the coefficient of determination; both indicate an exponential model is appropriate.

6. Press ⌨ and select Autoscale to see a scatter plot of the data.

7. Tap Menu to open the Plot view menu. Tap Fit to activate the fit. You will see the menu key change to Fit• to reflect that the fit is now active. The fit will be calculated and graphed in the same window, as shown to the right.

8. To see the expression for the fit, press ⌄ to move from tracing the scatter plot to tracing the fit, then tap Defn. In the figure, we restricted numeric display to 4 decimal places. Press Shift Settings (Home Settings) to view or change these settings.
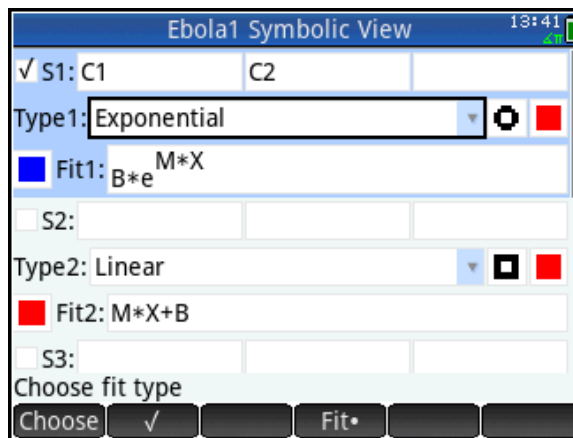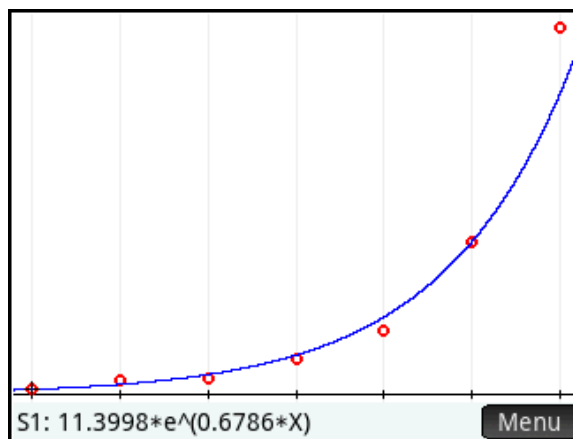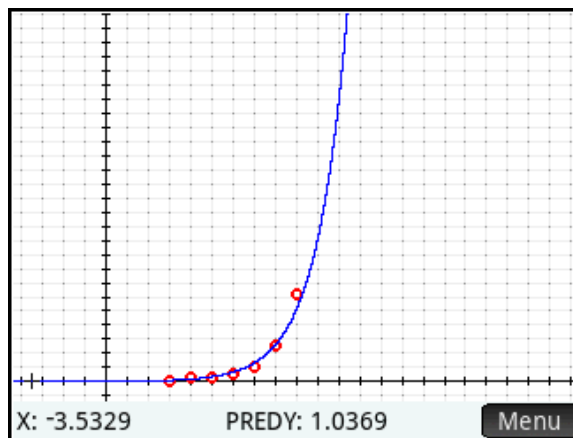
You can pinch to zoom horizontally, vertically, or square. Tap and drag to scroll the view. You can also use the app function *PredX()* to predict an x-value from a given y-value using the fit. There is also a corresponding *PredY()* app function to predict a y-value from a given x-value.

9. Trace to where the predicted value of y is close to 1. In the figure to the right, y is near 1 when x is approximately -3.5. What does this tell us about the first patient (Patient Zero)? Remember that x=3 represents March 24, 2014.

10. When does the model predict that there will be 1 million cases of Ebola in West Africa?

11. Press 🏠 to return to Home view. Press 🖩 to open the Toolbox menus. Tap ⏹App, tap Ebola1 (or Statistics 2Var), and select PredX. Enter 1 between the parentheses and press ⎵Enter to see the predicted value of x for y=1.

12. The figure to the right shows the predicted values of x for both y=1 and y=1,000,000.

Extension

In this extension, we transform the data to achieve linearity.

1. Press [Symb/Setup] to return to Symbolic view. Keep C1 as the independent data but make the dependent data LN(C2). For **Type1**, choose a `Linear` fit. Your Symbolic view should appear as in the figure to the right.
2. Again, press [View/Copy] and select `Autoscale` to see a scatter plot of the data.
3. Tap [Menu] to open the Plot view menu. Tap [Fit] to activate the fit.
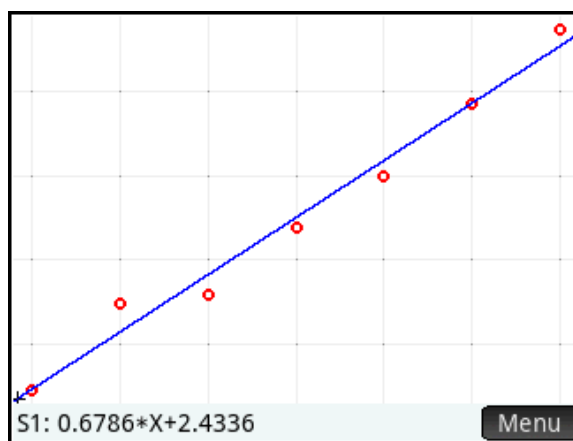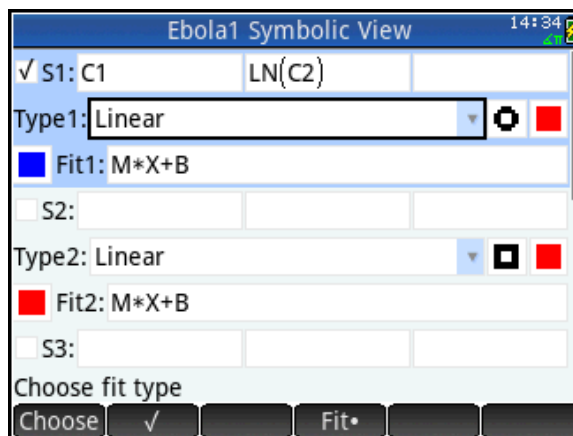
Of course you can always create a new column C3 and store LN(C2) into it, then plot C3 against C1. The advantage of the method shown here is that the Symbolic view retains the record of the independent and dependent data sets for any analysis you wish to perform. The definitions of each analysis appears explicitly in Symbolic view in S1–S5 and the data in Numeric view remains the original data.

Notice the same parameter, 0.6786 in both the linear and exponential expression. You can easily transform the equation LN(y)=0.6786*x+2.4336 to an exponential equation expressing y in terms of x using the CAS. Note that the CAS uses lowercase variable names. The steps are shown in the figure to the right.
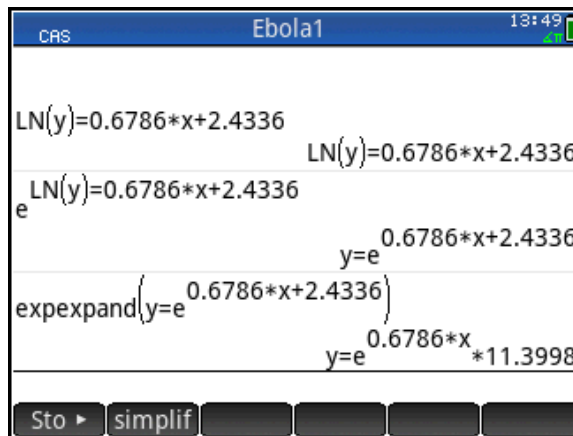
Part 2

From September 2014, efforts by the WHO and other organizations began to slow the spread of the disease noticeably. Monthly data through April 2015 are shown in the table to the right. The last data point represents the WHO Situation Report for April 26, 2015.

Remember that our exponential model predicted 1 million cases in May 2015. The data for the month before that date shows only 26,079 cases. In this second part, we examine the number of deaths using a different model.

1. Add the data from months 10-16 to your Statistics 2Var Numeric view, or get the Ebola2 app from your instructor. The Ebola2 app has the data already entered, as shown in the figure to the right.

The data for the months after Month 10 show a rapid slowdown in the spread of the virus. We now focus on the number of deaths and modeling these with a logistic model.

2. Press [Symb/Setup] to open Symbolic view. In S1, set C1 as the independent data and C3 as the dependent data. In **Type1**, select Logistic. Your Symbolic view should appear as in the figure to the right.

3. Look at the form of the logistic fit. As x increases, what value will the denominator approach? What value will the fit expression approach?

Ebola Cases and Deaths in 2014-2015: Month 3 = March 24, 2014		
Month	Cases	Deaths
3	86	59
4	242	147
5	270	183
6	599	338
7	1093	660
8	2599	1422
9	6242	2909
10	10114	4912
11	15319	5444
12	19463	7573
13	21724	8626
14	23694	9589
15	24907	10326
16	26079	10823

The value of L is known as the carrying capacity, the upper limit of the model.

4. What does L represent in our model?

5. Press [View Copy] and select `Autoscale` to see a scatter plot of the data.

6. Tap [Plot Setup] to open the Plot view menu. Tap [Fit] to activate the fit.

7. Press ⊙ to move from tracing the scatter plot to tracing the fit, then tap [Defn] to see the fit expression. How many deaths does the model predict by the end of the Ebola outbreak?

8. Use the model, or the PredY app function to predict the end of the outbreak. Why does the model make it difficult to do so?

9. When do you think the outbreak will end?

S1: 11,064.8186/(1+780.1603*e^(−0.616⁹   Menu
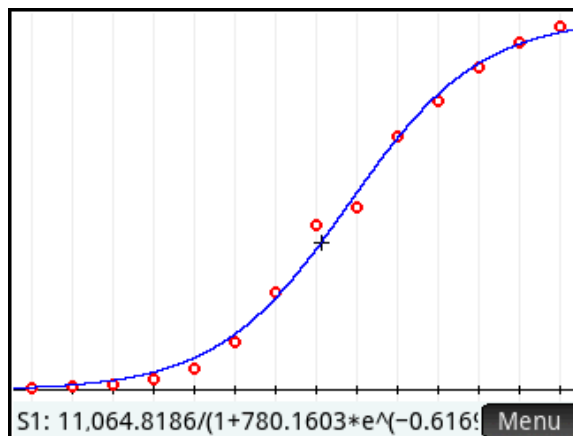
Answers

Part 1

4. There is no common difference between consecutive y-values. The number of cases more than doubles in some months and almost doubles in others. This add/multiply pattern of the x/y values indicate an exponential model.

9. X= -3.5 represents approximately September 10, 2014. In point of fact, Patient 0 was diagnosed early in December 2014.

10. Y=1,000,000 at approximately X=16.77; that x-value represents approximately May 18, 2015. The model shows why the WHO was so concerned about the Ebola outbreak and how quickly it could get out of hand.

Part 2

3. As x increases, the denominator approaches 1 and the y-values approach L.

4. L represents the total death toll from the outbreak.

7. The model predicts 11, 065 deaths from the outbreak.

8. The model approaches L as x approaches infinity.

9. As of May 2015, the three countries were reporting less than 20 new cases of Ebola per month. It is possible the outbreak will end as soon as August. Student answers may vary; the thought process in coming to an estimate is what matters here.

The data in this activity come from the World Health Organization, specifically their *Ebola Situation Reports*. The data can be found here:
http://www.who.int/csr/disease/ebola/situation-reports/archive/en/

Note that the exact day of the month varied slightly from month to month; we selected dates as close as possible to the 24th of each month.

Distributions and Sampling

In this multi-part activity, we examine various plots of univariate data, as well as numerous ways of generating such data. We then look at sampling distributions and how to generate them.

HP Prime Functionality Introduced:

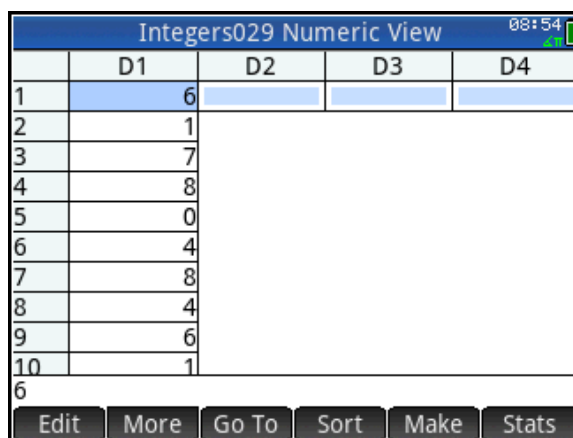Using the Statistics 1Var app Numeric, Symbolic, and Plot views; the Make Column Data wizard, Copy and Paste

AP Statistics Content:

Constructing and interpreting graphical displays of univariate data; summarizing distributions of univariate data; the normal distribution; discrete random variables and their probability distributions, including binomial and geometric; simulation of random behavior and probability distributions; simulation of sampling distributions and the Central Limit Theorem

Activity 1: Generating Data

Your instructor will send you an app named *Integers029*.

1. Press ![Apps] to open the App Library. Locate the *Integers029* app and either tap it or select it and tap ![Start].

2. This app is a copy of the Statistics 1Var app and has all of that app's functionality and views. The app opens in Numeric view with 50 random integers from 0 to 9 in list D1.


In Numeric view, the menu keys are:

- ![Edit]: edit the current cell
- ![More]: opens a menu with options to select or delete multiple cells
- ![Go To]: jump to a particular cell
- ![Sort]: sort column data
- ![Make]: create a list of data
- ![Stats]: display summary statistics

3. Press ![Symb Setup] to open Symbolic view, which contains 5 analyses, named H1 through H5. Each analysis can use one list for its data, with another (optional) list for the frequencies. You can choose a plot type for each analysis, and a color for the plot. Set H1 to use D1 as its data and to create a dot plot in Plot view, as shown in the figure to the right. The menu keys here are:

- [Choose]: make a selection from a list of options
- [√]: check to activate (or uncheck to de-activate) H1-H5
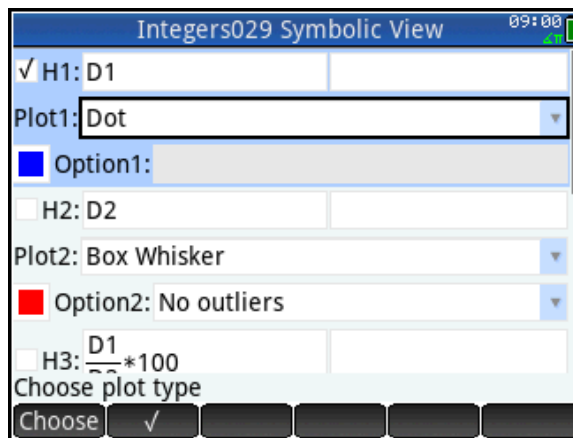
4. Press ![View Copy] and select Autoscale to see the dot plot. Pinch horizontally to make the dots circular if they are distorted.

The tracer indicates we are tracing H1, the current x-value is X=0, and the frequency of 0 in the data set is F=5. Tap anywhere to move the cursor to that x-value, or use the direction pad to move the tracer.
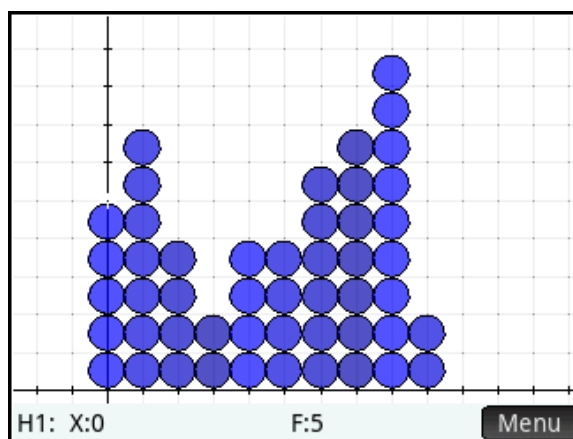
We now look at how we generated those 50 integers from 0 to 9 in the first place.

Press [Num] to return to Numeric view. Press [Shift] [Esc] (Clear) and a pop-up box will ask what you want to delete. Tap All to clear Numeric view.

5. Tap [Make] to open the **Make Column Data** wizard. By default, the **Expression** field is highlighted. Press [Mem B] to open the Toolbox menus and tap [Math]. Tap Probability, then Random, and select Integer. Between the parentheses, enter 0,9 to complete the command. In the **Stop** field, enter 50. You can also choose where to save the data generated via the **Col** (Column) field. The figure to the right shows the completed wizard.

6. Tap [OK] to see the data generated and stored in list D1 in Numeric view. Pinch vertically to change the font size, if you wish. In the figure to the right, we pinched vertically to choose a smaller font size.
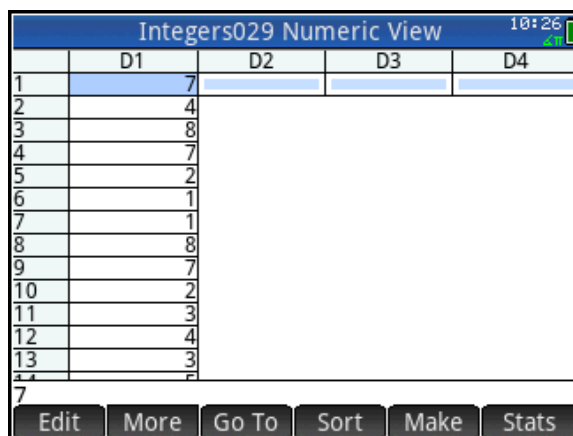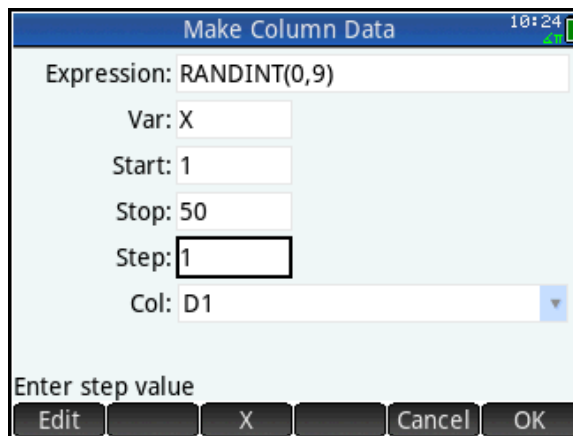
Suppose we wanted 100 random integers from a normal distribution with a mean of 5 and a standard deviation of 2.

7. Tap [Make] to return to the **Make Column Data** wizard. In the Expression field, press [Mem B] to open the Toolbox menus and tap [Math]. Tap Numbers and select Integer Part. Between the parentheses, press [Mem B] and tap [Math]. Tap `Probability`, then `Random`, and select `Normal`. Between the parentheses, enter 5,2. In the **Stop** field, enter 100. The completed wizard is shown to the right.

8. Tap [OK] to see the data generated and stored in list D1 in Numeric view.

9. Press [Symb Setup] to return to Symbolic view. Set H1 to use D1 for its data and to plot a histogram. Press [View Copy] and select `Autoscale` to see the histogram.
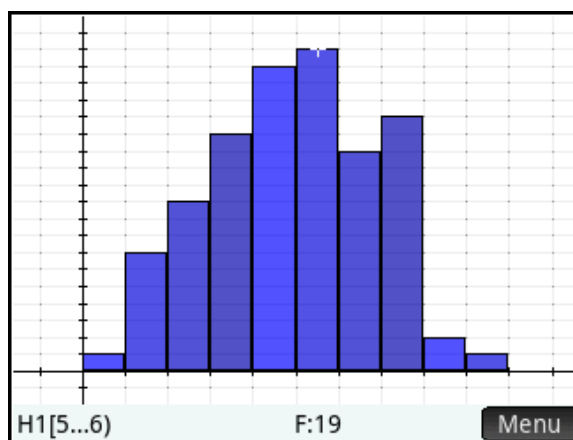
The figures below show some other data as generated via `Make`. On the left is the completed wizard and on the right is the histogram of the data. The random commands used here can all be found by pressing [Units], tapping `Catlg`, and pressing [8 R] to jump to commands that start with the letter R. Scroll down until you see commands that start with "rand". Besides the random integer and random normal commands, you will see randbinomial, randchisquare, randexp, randfisher, and randgeometric. Scroll further to see randpoisson and randstudent.
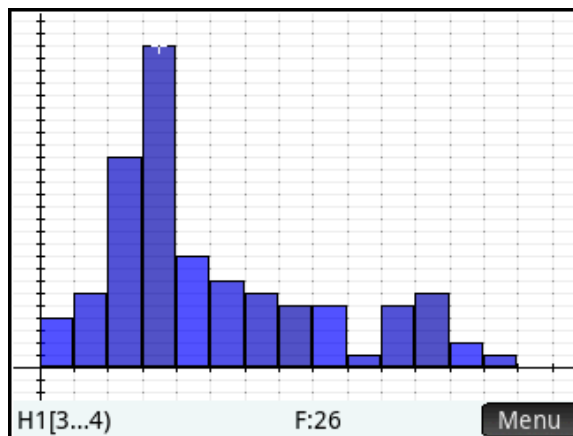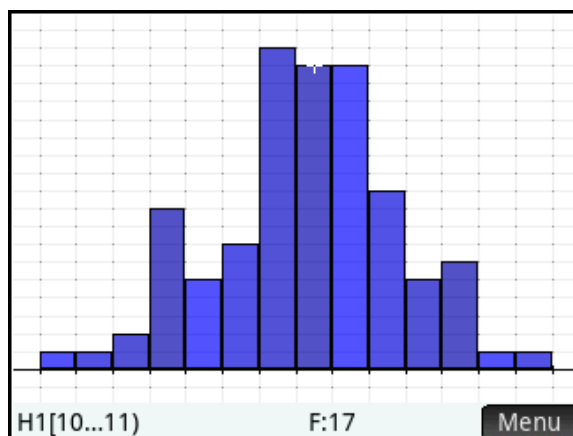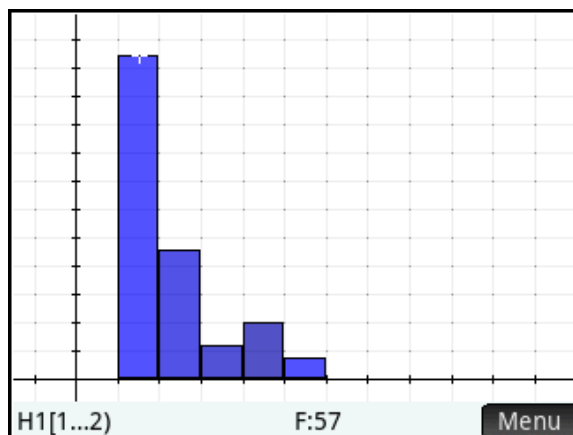
Activity 2: Sampling Distributions

The previous technique can be extended to produce sampling distributions. Suppose that a bag contains 200 marbles, 100 yellow and 100 green. We want to simulate choosing samples of size n=20 from that bag and computing the percentage of green marbles in each sample. We want to see the sampling distribution of the proportion of green marbles drawn after 500 simulations.

1. First, we illustrate the process in Home view. The randbinomial() command we know can be found in the Catalog menu. The others are all found by pressing [Mem B], then tapping [Math] and selecting `List`.

*   randbinomial(1, 0.5) returns either 0 or 1 with equal probability. This simulates drawing one marble, where we define 0 to mean a yellow marble was drawn and 1 to mean a green marble was drawn.

*   MAKELIST(randbinomial(1, 0.5), A, 1, 20)) creates a list of 20 draws.

*   ΣLIST(MAKELIST(randbinomial(1, 0.5), A, 1, 20))) totals all the ones; in other words, it counts up the total number of green marbles in the 20 draws.

*   The final step divides by 20 to get the percentage of green marbles drawn.

2. Return to Numeric view and tap [Make]. In the **Expression** field, press [Shift] [Menu Paste] (Paste) and select the last expression we entered in Home view. In the **Stop** field, enter 500 to run the simulation 500 times. Tap [OK] to see the proportion of green marbles drawn in each of the 500 simulations.

3. Press ⎇Shift ⎇Plot to open Plot Setup.
   Set both **H Width** (the bin width) and
   X Tick to 0.05. Press ⎇Symb to open
   Symbolic view and set **H1** to use D1
   for its data and to draw a histogram.

4. Press ⎇View and select `Autoscale`
   to see the histogram.

As expected, the sampling distribution of
the proportion of green marbles drawn is
approximately normal, with a mean near
0.5.
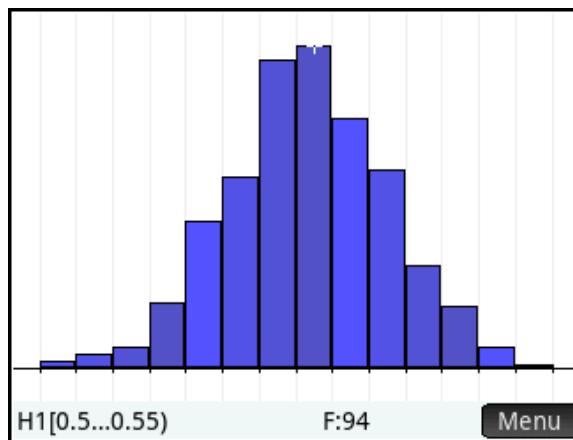
500 samples of size n=20

5. You can return to Numeric view, tap
   `Make` and re-paste the expression
   into the **Expression** field. Now you
   can change the sample size from 20
   to any other value. You can also
   change the number of samples
   using the **Stop** field. The rest of the
   figures on this page show different
   numbers of samples of various
   sizes.
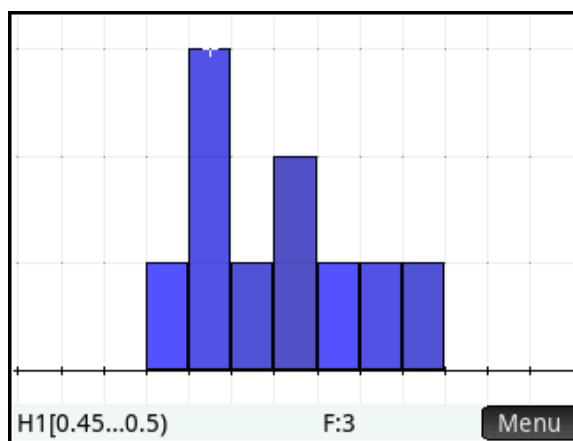
10 samples of size n=20
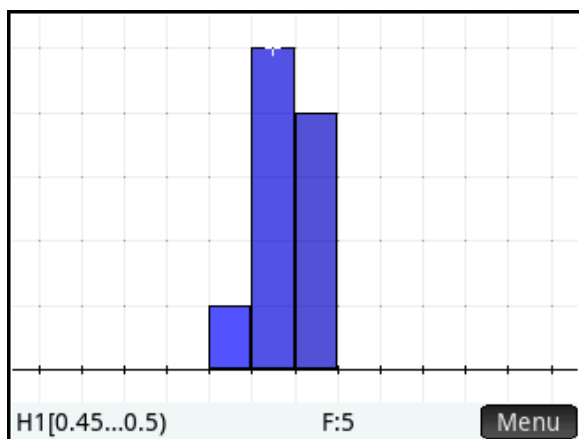
10 samples of size n=200
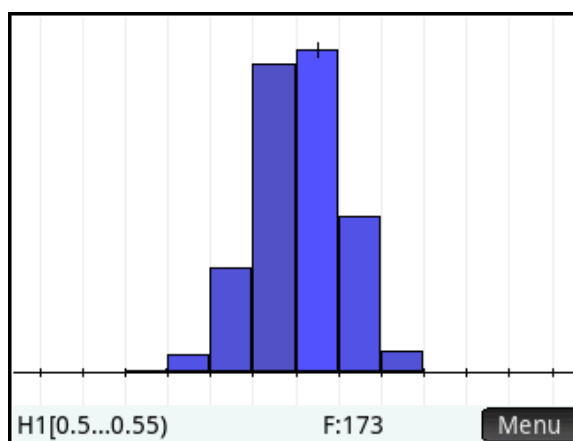
500 samples of size n=100

In the next example, we look at the sampling distribution of a mean.

Activity 3: The Central Limit Theorem

In this activity, we look at the sampling distribution of means from a population with a pronounced skew.

The figure to the right shows a sampling distribution created using the randgeometric(0.5) command to generate 500 numbers from a geometric distribution in which the probability of success is 0.5. The distribution is strongly skewed. As you can see from the histogram, the mean of the sampling distribution must be close to 2. What does the sampling distribution of the means look like if we repeated this simulation 500 times?
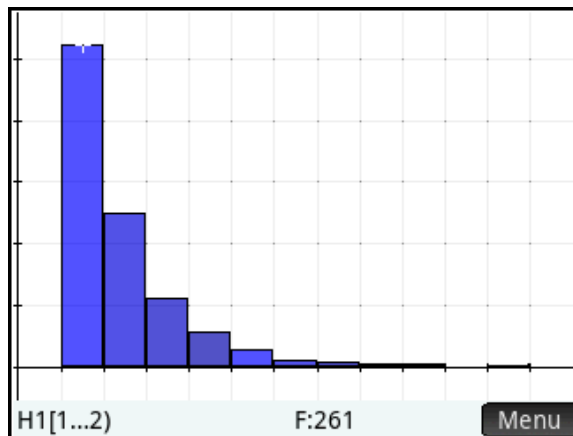
1. Again, we start in Home view, testing our procedure step-by-step:

- Randgeometric(0.5) returns an integer that represents the number of trials needed before a success if the probability of success is 0.5.

- MAKELIST(randgeometric(0.5), A, 1, 500) returns a list of 500 trials.

- ΣList(MAKELIST(randgeometric(0.5), A, 1, 500))/500 returns the mean of the sample of size n=500.

2. Return to Numeric view and tap Make. In the **Expression** field, press Shift ☰Menu/Paste (Paste) and select the last expression we entered in Home view. In the **Stop** field, enter 100 to run the simulation 100 times. Tap OK to see the sample of 100 means in list D1.

3. Press **Shift** **Plot↗** to open Plot Setup.
    Set both **H Width** (the bin width) and
    X Tick to 0.05. Press **Symb↗** to open
    Symbolic view and set **H1** to use D1
    for its data and to draw a histogram.
4. Press **View** and select `Autoscale`
    to see the histogram.

As shown in the figure to the right, the
sample distribution of the means is both
symmetric and unimodal; that is to say, the
distribution of the means is normal, with a
mean quite close to 2.

You can return to Numeric view, re-paste
the expression in the **Make Column Data**
wizard, and change the sample size from
500 to any other positive integer value. You
can also change the number of samples
using the **Stop** field. The rest of the figures
on this page show different numbers of
samples of various sizes.
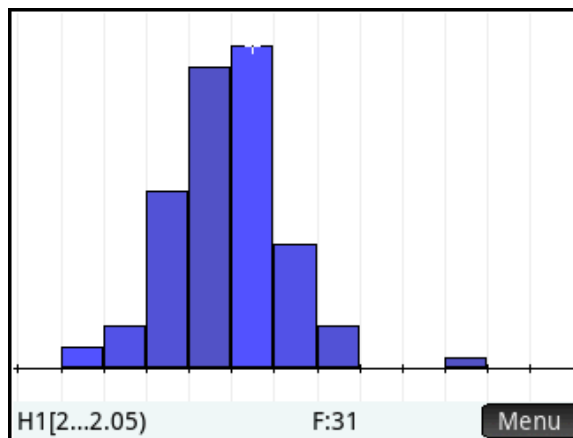
100 samples of size n=500
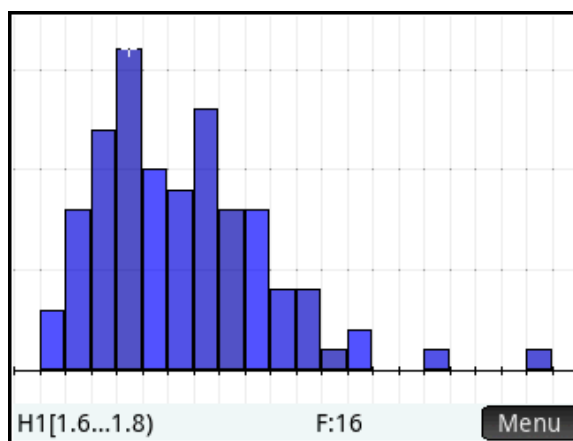
100 samples of size n=5
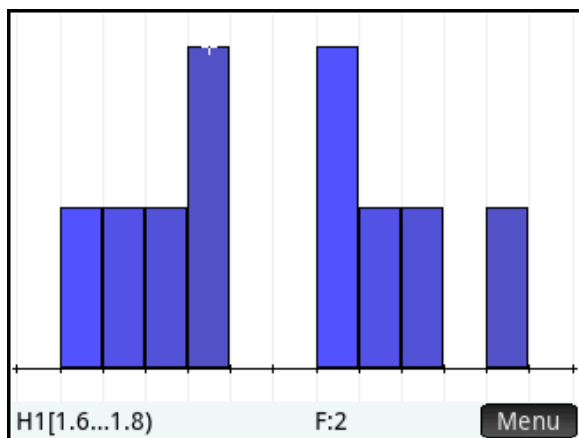
10 samples of size n=5

Use the *Make Column Data* wizard to generate univariate data and to simulate sampling
distributions of means and proportions. Test your procedure step-by-step in Home view and
paste the final expression into the wizard. Choose the number of samples in the **Stop** field.
Change the number of samples and the sample size to explore the Central Limit Theorem!

Discrete Random Variables and Their Distributions

HP Prime Functionality Introduced:

Using the Statistic 1Var app Numeric, Symbolic, and Plot views; the BINOMIAL(), BINOMIAL_CDF(), and BINOMIAL_ICDF() functions; the GEOMETRIC(), GEOMETRIC_CDF(), and GEOMETRIC_ICDF() function

AP Statistics Content:

Discrete random variables and their distributions, especially the binomial and geometric distributions

Activity 1: A binomial distribution

What is the theoretical distribution of the number of heads obtained when a fair coin is tossed 10 times?

To create a histogram of the theoretical distribution, we will make a column of data that represents the possible numbers of heads (0-10), and another column of data that represents the probabilities associated with each number of heads.

1. Press $\boxed{\text{Apps Info}}$; then reset and start the *Statistics 1Var* app.
2. In Numeric view, tap $\boxed{\text{Make}}$. To make a list of the integers from 0 to 10 in D1, just change the value in the **Start** field to 0 and tap $\boxed{\text{OK}}$. You will see the integers from 0 to 10 in D1.
3. Move the cursor to D2 and tap $\boxed{\text{Make}}$ again. Press $\boxed{\text{Mem B}}$, tap $\boxed{\text{Math}}$, tap `Probability`, `Density`, and select `Binomial`. Between the parentheses, enter 10, 0.5, X, as shown in the figure to the right. Again, change the value in the **Start** field to 0. Tap $\boxed{\text{OK}}$ to see the probabilities in D2 (next page).
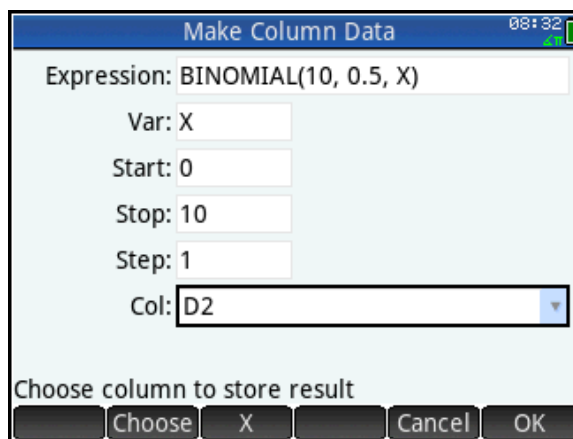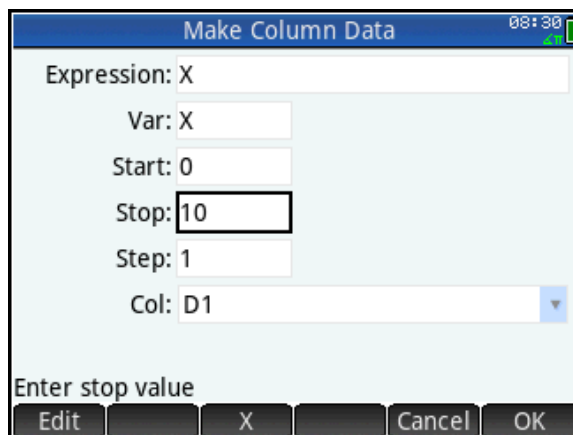
As you can see from the figure, the distribution of the probabilities is exactly symmetric, with 5 heads having the greatest probability.

	D1	D2	D3	D4
1	0	0.0009765		
2	1	0.0097656		
3	2	0.0439453		
4	3	0.1171875		
5	4	0.2050781		
6	5	0.2460937		
7	6	0.2050781		
8	7	0.1171875		
9	8	0.0439453		
10	9	0.0097656		
0				

Statistics 1Var Numeric View — Edit | More | Go To | Sort | Make | Stats

4. Press ⬚ to open Symbolic view and set H1 to use D1 for the data and D2 for the frequencies. Set **Plot1** to Histogram.

Statistics 1Var Symbolic View
√ H1: D1   D2
Plot1: Histogram
☐ Option1:
☐ H2:
Plot2: Histogram
☐ Option2:
☐ H3:
Choose plot type
Choose | √

5. Press ⬚ and select Autoscale to see the histogram.

H1[5...6]   F:0.24609375   Menu

6. What is the total area of this histogram? To answer this question, go to Home view and enter

    ΣLIST(D2), as shown to the right.

7. What is the probability that less than 5 heads are obtained in a set of 10 tosses? The figure shows two ways to approach this question.
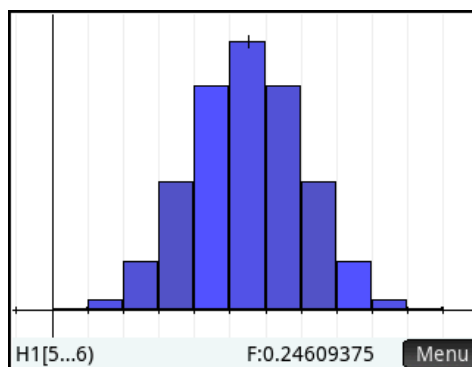
Statistics 1Var

ΣLIST(D2)   1
D2(1)+D2(2)+D2(3)+D2(4)+D2(5)
                                                  0.376953125
BINOMIAL_CDF(10,0.5,4)   0.376953125
Sto ►

Activity: A Geometric Distribution

A geometric distribution is usually (but not always) defined as the probability distribution of the number of trials needed to get one success, based on the probability of success. These probabilities form a geometric sequence, hence the name of the distributions. What is the theoretical distribution of the probability of a fair coin landing on heads on the kth trial?

We proceed much as we did for the binomial distribution to generate the theoretical probabilities associated with the first 10 trials.

1. Reset and open the Statistics 1Var app. Use `Make` to put the integers 1-10 in D1 (not 0-10).
2. In Numeric view, place the cursor in D2 and tap `Make` again. In the Expression field, enter GEOMETRIC(0.5, X), as shown in the figure.
3. Tap `OK` to see the probabilities in D2.
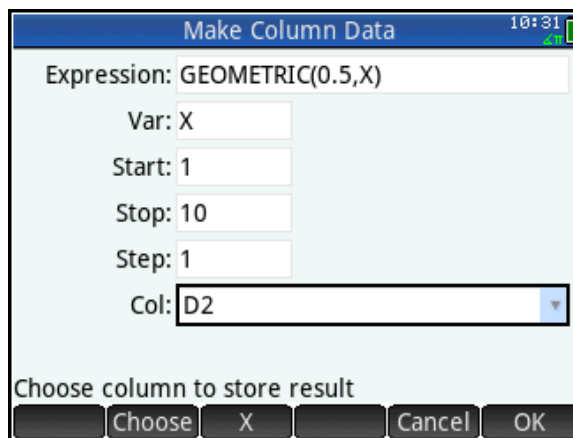
4. Press `Symb` to open Symbolic view and set H1 to use D1 for the data and D2 for the frequencies. Set **Plot1** to `Histogram`.
5. Press `View` and select `Autoscale` to see the histogram.
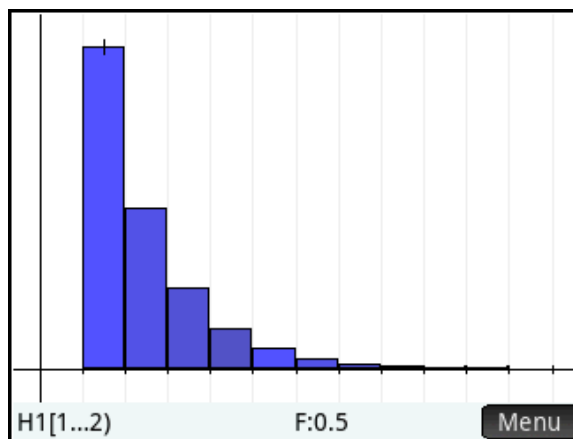
Teacher Notes

Binomial distributions

The binomial probability mass function is given by $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$, where

$\binom{n}{k} = \dfrac{n!}{k1 \cdot (n-k)!}$ or $_nC_k$ (COMB(n,k)). If you would like to start with this definition and show that

it agrees with the built-in BINOMIAL() function, you can use this definition with [Make] or use the Spreadsheet app, as shown in the following example.

1. Press [Apps Info] and select Spreadsheet. Tap on the header for Column A and enter =Row-1. When you press [Enter ≈], Column A will fill with the non-negative integers.

The system variable *Row* represents the current row. So in A1, Row=1 and Row-1=0.

2. Now tap on the header for Column B and enter our formula, as shown to the right. Here we use n=10 and k=Row-1. Press [Enter ≈] to see the probabilities associated with each of the positive integers.

One benefit of this approach is that it clearly shows that the probability for k>n is zero (see the figure below right).

3. Tap and hold on cell A1. You will see the menu key [Select] change to [Sel•]. Now drag to cell B11 to select the data we want. Press [Shift] [▲View Copy] (Copy) to copy the data to the clipboard.

4. Reset and open the Statistics 1Var app. With the cursor in D1, press [Shift] [≡Menu Paste] (Paste) and select `Grid data` to paste the data into the Statistics 1Var app and proceed from Step 4 in the activity.

	Spreadsheet			09:39
hp A	B	C	D	E
1 0				
2 1				
3 2				
4 3				
5 4				
6 5				
7 6				
8 7				
9 8				
10 9				
=Row-1				

Edit | Format | Go To | Select | Go ↓

	Spreadsheet			09:29
hp A	B	C	D	E
1 0	0.0009765625			
2 1	0.009765625			
3 2	0.0439453125			
4 3	0.1171875			
5 4	0.205078125			
6 5	0.24609375			
7 6	0.205078125			
8 7	0.1171875			
9 8	0.0439453125			
10 9	0.009765625			
=COMB(10,Row-1)*0.5^(Row-1)*0.5^(10-(Row-1))				

Edit | Format | Go To | Select | Go ↓

	Spreadsheet			09:31
hp A	B	C	D	E
6 5	0.24609375			
7 6	0.205078125			
8 7	0.1171875			
9 8	0.0439453125			
10 9	0.009765625			
11 10	0.0009765625			
12 11	0			
13 12	0			
14 13	0			
15 14	0			
=COMB(10,Row-1)*0.5^(Row-1)*0.5^(10-(Row-1))				

Edit | Format | Go To | Select | Go ↓

Geometric distributions

For geometric distributions, the probability function is defined as $P(X=k)=p*(1-p)^{k-1}$.

A similar approach can generate as much of a geometric distribution as you care to copy in the Spreadsheet app. The difference is that here the domain is the positive integers, so we start with X=1.

Of course, you can also use [Make] directly in the Statistics 1Var app.
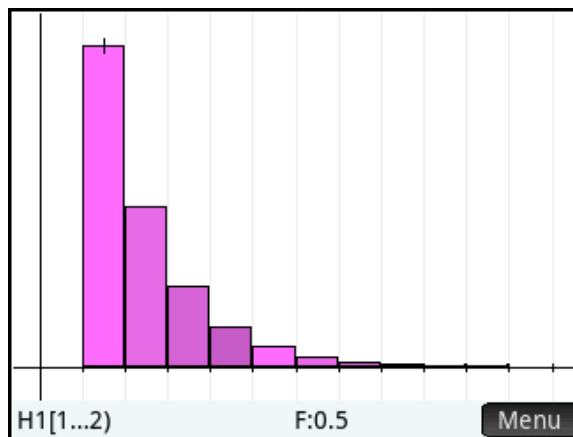
In this case ΣLIST(D2) will return a value close to but less than one. What is the probability of heads in the first 4 tosses? Two ways to approach this question are shown to the right.

The Normal Distribution

This activity explores the normal distribution, the normal probability density function, the normal cumulative probability density function, and the normal inverse cumulative probability density function.

HP Prime Functionality Introduced:

Using the Solve app Numeric, Symbolic, and Plot views; the NORMALD(), NORMAL_CDF(), and NORMALD_ICDF() functions

AP Statistics Content:

The normal distribution; properties of the normal distribution

Activity 1: The Normal Probability Density Function

NORMALD(μ, σ, x) computes the probability density at x for the normal distribution with mean μ and standard deviation σ.

1. Open the Function app: press [Apps Info] and select Function. The Function app opens in Symbolic view, where you can enter up to 10 function expressions in X. In F1(X), enter NORMALD(0,10,X) as shown in the figure to the right. Remember that the normal probability density function is found by pressing [Mem B], then tapping Probability, Density, and selecting Normal.

2. Press [Plot Setup] to see the graph. Press [+ Ans ;] once or twice to zoom in until the domain is approximately [-4, 4]. Pinch vertically until the maximum of the curve is near the top of the display. Use the figure to the right as a guide.
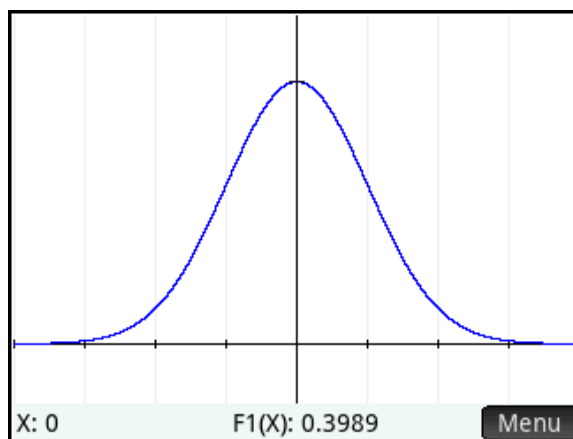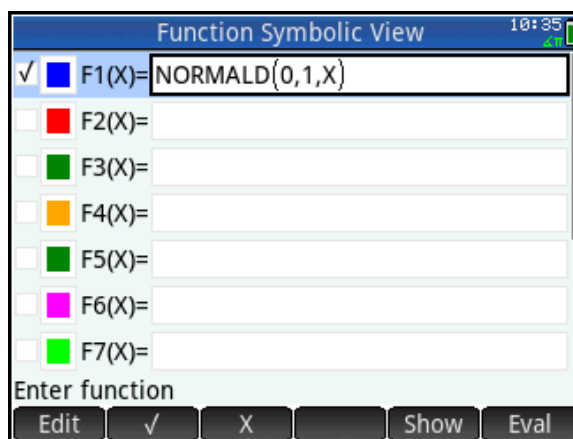
3. We wish to estimate the area under the graph; that is, the area between the curve and the x-axis. On the figure from the previous page, draw the triangle whose base has endpoints at (-3,0) and (3,0) and whose third vertex is the maximum of the graph. What is the area of this triangle? Do you think this area over- or under-estimates the true area under the curve?

    _____

    _____
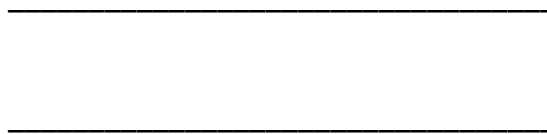
4. Tap ⬚Menu⬚ to open the Plot view menu. Tap ⬚Fcn⬚ and select Signed area…. You will be prompted for a lower endpoint for computing the area. Type -4 and press ⬚Enter⬚. Next you will be prompted for an upper endpoint for the area. Type 4 and press ⬚Enter⬚.

5. What do you think is the total area under this curve?



Signed area: 0.9999   OK

_____
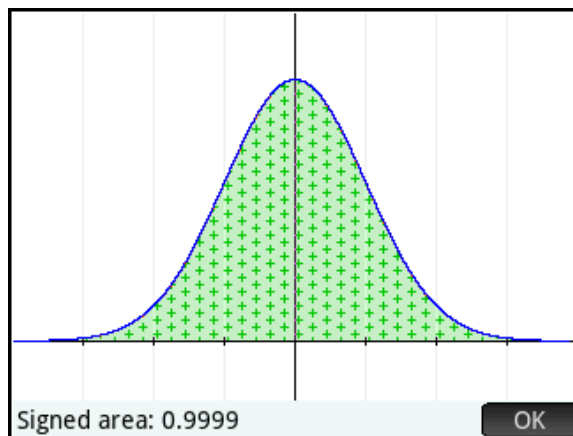
The area under this curve, since it is exactly one, can be interpreted as a probability. In this case, we call the curve a probability density function; more precisely, this one is the standard normal probability density function. There are three areas that are important to us:

*   The area to the left of an x-value (called the lower tail) represents the probability that a randomly-chosen x-value is no greater than the x-value
*   The area to the right of an x-value (called the upper tail), represents the probability that a randomly chosen x-value is no less than the x-value
*   The area between two x-values represents the probability that a randomly chosen x-value is between these two x-values

In this activity, to calculate a lower-tail probability, we use a number such as -10 for the lower endpoint as a substitute for $-\infty$. Similarly, to calculate an upper-tail probability, we use a number such as 10 for the upper endpoint as a substitute for ∞.

For example, to find the probability that a randomly chosen x-value is 3 or less, we repeated the steps in #4 above with a lower endpoint of -10 and an upper endpoint of 3. The area means the probability is P=0.9987 that a randomly chosen x-value is 3 or less. Note that this also means the probability of a randomly chosen x-value being greater than 3 is P=1-0.9987 or P=0.0013.

Signed area: 0.9987

6. What is the probability that a randomly chosen x-value is:
   a.   -2 or less?
    b. 1 or more?
    c. Between -1 and 1?
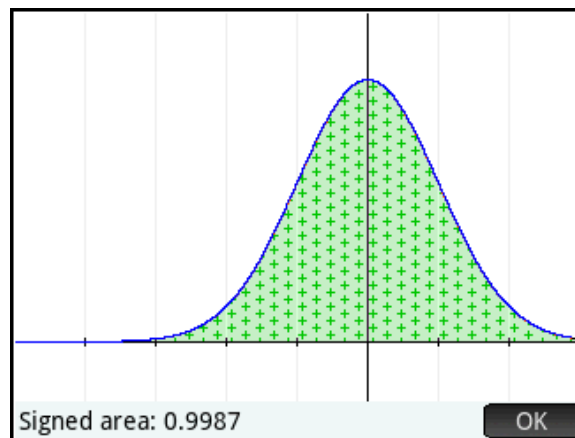
We will now define F2(X) to be the area function for our normal probability density function.

7. Press ⬚ to return to Symbolic view. In F2(X), enter AREA(F2, -10, X) as shown to the right. To find the AREA() function, press ⬚, tap App, tap Function and select AREA. To enter F, press ⬚ ⬚. To enter X, use the menu key.
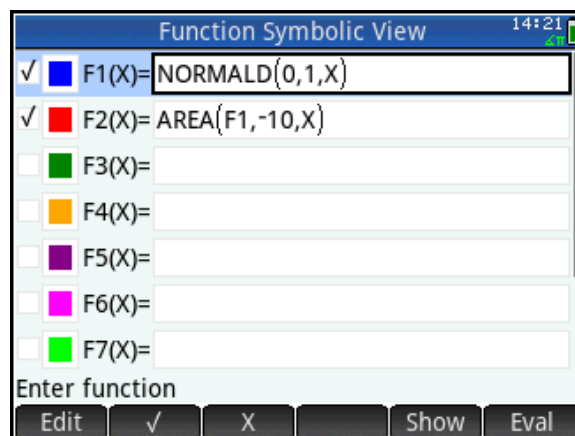
8. Press ⬚ to view the graph. Pinch to zoom and drag to scroll until both functions are visible and the domain is roughly [-4, 4], as shown in the figure.
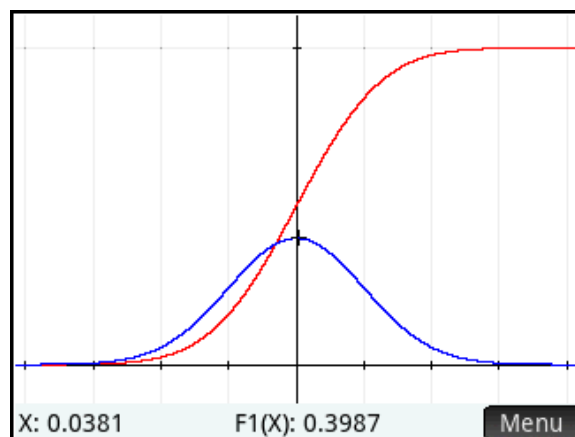
9. What are the equations of the horizontal asymptotes of F2(X)? Explain why these would be expected from a function that reports the area of F1(X).

    _____

    _____

10. Given the symmetry of F1(X), what do you expect the value of F2(0) to be? Press ⊙ to move from tracing F1(X) to tracing F2(X) and enter 0. Is the tracer value as you expected?

    _____

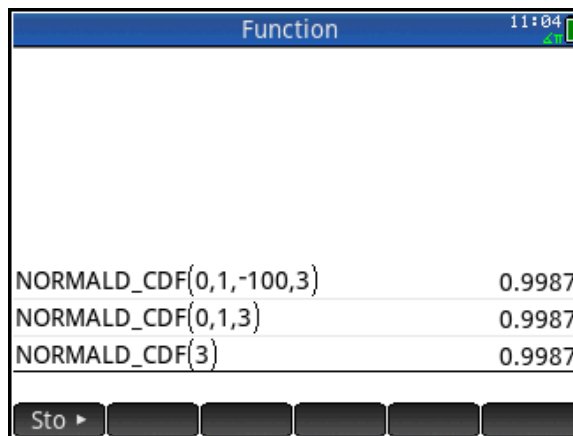11. To find the probability that a randomly chosen x-value from a normal distribution is less than 3, simply compute the area between -10 and 3 in Plot view or F2(3) in Home view. This calculation and its result are shown in the figure to the right, along with the corresponding calculations from #6 above.

Function	14:20
F2(3)	0.9987
F2(-2)	0.0228
F2(100)−F2(1)	0.1587
F2(1)−F2(-1)	0.6827

Sto ▶

Activity 2: Using the Normal Probability Density Functions

In the previous activity, you used an area function to calculate probabilities associated with the normal probability density function. The cumulative normal probability density function, NORMAL_CDF(), can be used to find these same probabilities directly. NORMAL_CDF(µ, σ, *lower*, *upper*), returns the probability that a randomly chosen *x*-value from a normal distribution with mean µ and standard deviation σ will be between *x=lower* and *x=upper*. In Home view, let's repeat the last four calculations.
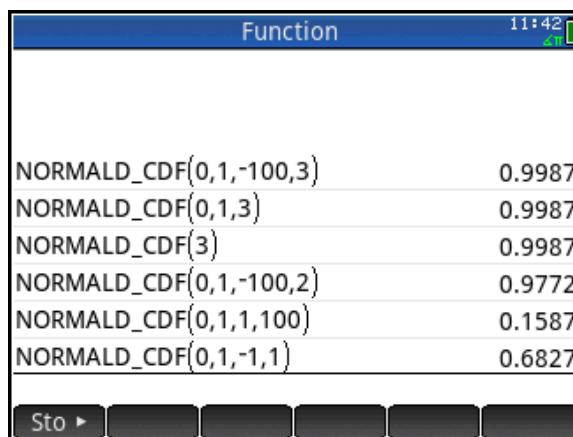
1. Press ⌂ to open Home view. Press, tap Math, tap Probability, tap Cumulative, and select Normal. Between the parentheses, type 0, 1, -100, 3 and press Enter. The input and output is shown in the figure to the right. The p-value returned agrees with our previous area calculation.

NORMALD_CDF(0,1,-100,3)	0.9987
NORMALD_CDF(0,1,3)	0.9987
NORMALD_CDF(3)	0.9987

If only one bound is given, it will be assumed as the upper bound and the lower-tail probability will be returned. Thus NORMAL_CDF(0, 1, 3) returns the same result. In fact, in this last case, if µ=0 and σ =1, they may be omitted. So NORMALD_CDF(3) also returns the same result, the lower-tail probability for x=3.

NORMALD_CDF(0,1,-100,3)	0.9987
NORMALD_CDF(0,1,3)	0.9987
NORMALD_CDF(3)	0.9987
NORMALD_CDF(0,1,-100,2)	0.9772
NORMALD_CDF(0,1,1,100)	0.1587
NORMALD_CDF(0,1,-1,1)	0.6827

2. Repeat the calculations from #6 using the NORMAL_CDF() function.

The lower-tail probabilities present an interesting case. Since there is a 1:1 correspondence between the x-values and their lower-tail probabilities, we can find the x-value associated with any lower-tail probability. This new function is called, appropriately, the inverse cumulative normal probability density function. NORMAL_ICDF(µ, σ, *p*) returns the *x*-value associated with the lower-tail probability *p* of the normal distribution with mean µ and standard deviation σ.

Where is the first quartile of the standard normal distribution? In other words, what x-value has a lower-tail probability of p=0.25?

3. Press ⌂ (Settings) to open Home view. Press (Mem B), tap Math, tap Probability, tap Inverse, and select Normal. Between the parentheses, type 0, 1, 0.25 and press Enter. The input and output is shown in the figure to the right.
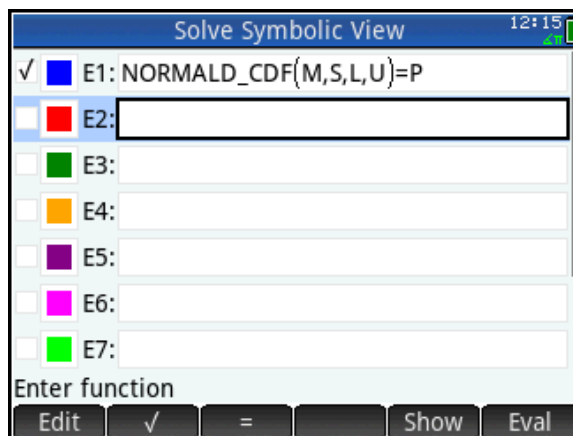
Like the NORMAL_CDF() function, the mean and standard deviation can be omitted for the standard normal distribution. Thus NORMALD_ICDF(0.25) returns the same result. The figure above contains both variations, as well as a check using NORMAL_CDF() to calculate the lower-tail probability of x= -0.6745. The check conforms to our expectations.

The HP Prime Solve app gives us a method for solving problems of this type with a unique and easy to use interface.

4. Press Apps (Info) and select Solve. The Solve app opens in Symbolic view, where you can enter up to 10 equations. In E1, enter NORMALD_CDF(M, S, L, U) = P, as shown to the right. The variables are as follows:

- M: the mean of the distribution
- S: the standard deviation
- L: the lower bound
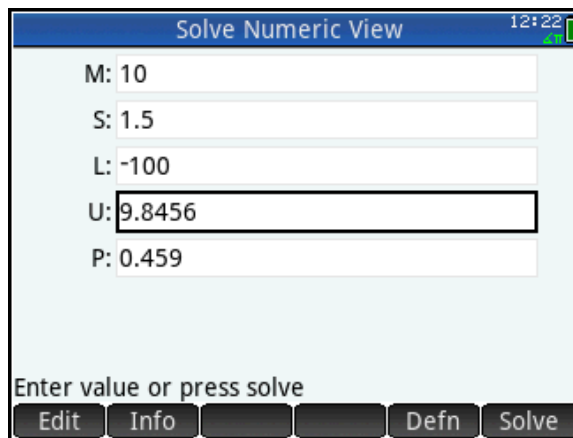- U: the upper bound
- P: the probability

Example 1: In a normal distribution with μ=10 and σ=1.5, what x-value has a lower-tail probability of p=0.459?

5. Press 🔲 to open Numeric view. Enter M=10, S=1.5, L=-100, and P=0.459, as shown in the figure.
6. Highlight **U** and tap ⎡Solve⎤ to see U=9.8456

Example 2: In a normal distribution with μ=5 and σ=0.5, what is the probability that a randomly-chosen x-value is at least 4?

1. Enter M=5, S=0.5, L=4, and U=100, as shown in the figure.
2. Highlight **P** and tap ⎡Solve⎤ to see P=0.9772
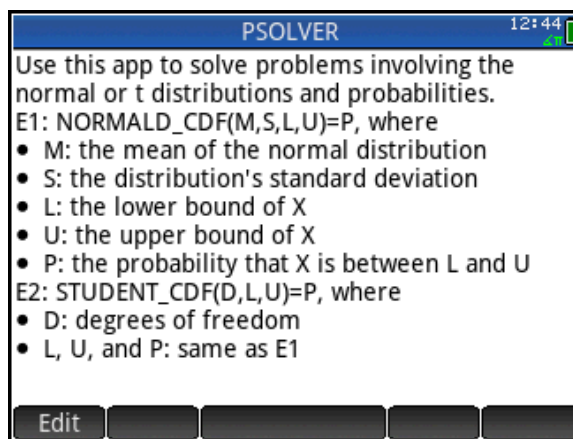
In general, the Solve app is quite useful in creating solvers for specific types of problems. You can add a note that contains the definitions of the variables and a brief description of each equation.

3. Press ⎡Shift⎤ ⎡Apps⎤ and tap ⎡Edit⎤ to create your note.
4. When you are done, press ⎡Apps⎤ and tap ⎡Save⎤ to save your app with a new name. In the figure to the right, our new app appears in the App Library as *PSOLVER*.

When you send PSOLVER to your students, the note and the equation are both included. Later, you can add an equation in E2 to cover the t-distributions.

Answers

Activity 1

3. The area of the triangle is 0.5*6*0.3989 or 1.1967. This appears to be an over-estimate, though the extreme lengths of the tails in this case may give students difficulty in estimating their areas.

5. The area under the curve looks as though it could be exactly 1.

6.
a. The probability that x is 2 or less is 0.9722.
b. The probability that x is 1 or more is 0.1587.
c. The probability that x is between -1 and 1 is 0.6827.

9. The asymptotes are y=0 and y=1. This assures us that 0<y<1, which is appropriate for probabilities.

10. We would expect F2(0)=0.5 and it does.

Teacher Notes

All of the HP Prime cumulative probability density functions behave the same way as the cumulative normal. With one x-value, they return the lower-tail probability associated with that x-value for the defined distribution (binomial, geometric, t, X^2, etc.). With two x-values, they return the probability that a randomly chosen x-value will lie between those two x-values for the defined distribution. The figure to the right illustrates some examples.

Confidence Intervals

In this activity, we create confidence intervals in a variety of contexts, using HP Prime.

HP Prime Functionality Introduced:

Using the Inference app Numeric, Symbolic, and Plot views

AP Statistics Content:

Large sample confidence interval for a proportion, a mean, a difference between two means, and a difference between two proportions

Problem 1

A coin is tossed 30 times and comes up heads 11 times. Construct and interpret a 95% confidence interval for the proportion of heads from this coin.

First, $\hat{p} = \dfrac{11}{30} \approx 0.3667$ and $1 - \hat{p} \approx 0.6333$.

Second, z*=NORMALD_ICDF(0.025)≈1.96

So the 95% confidence interval is given by:

$$\hat{p} \pm 1.96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$0.3667 \pm 1.96 \cdot \sqrt{\frac{.3667 \cdot (1 - 0.3667)}{30}}$$

The 95% confidence interval is thus

(0.1942, 0.5391). The steps and final answer is shown in the figure. Note the extensive use of Copy and Paste to make these calculations less laborious.

1. Interpret the 95% confidence interval. Specifically, do you have any reason to believe that the coin is biased?

_____

_____

Now let's use the Inference app.

2. Press **Apps** and select `Inference`. The app opens in Symbolic view. For **Method**, select `Confidence interval` and for **Type**, select `Z-Int: 1 π`, as shown in the figure.

3. Press **Num** to open Numeric view. From our problem, enter x=11, n=30, and C=0.95, as shown in the figure.

4. Tap `Calc` to see the confidence interval displayed along with the confidence level and the critical value of Z. The confidence interval shown here agrees with our previous calculations.

5. Tap `OK` to return to Numeric view.

6. Press **Plot** to open Plot view. Here the confidence interval is presented graphically. The display clearly shows the mapping between our random variable x and the normal distribution z.

7. Tap `C`; a white dot will appear to show the option is active. Press ▲ and ▼ to increase and decrease the value of C, dynamically illustrating the relationship between the magnitude of C and the width of the confidence interval.

8. Tap `C` again to de-activate the option.
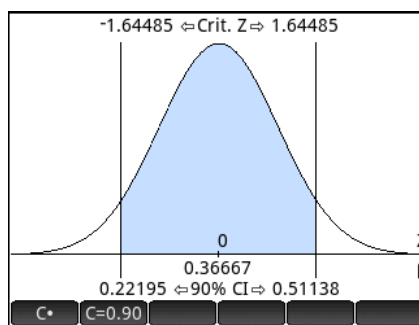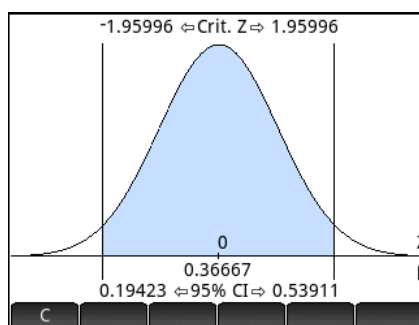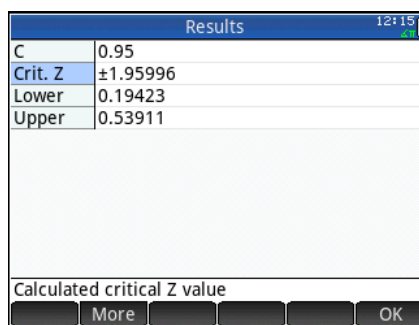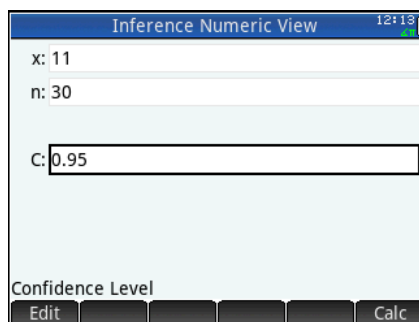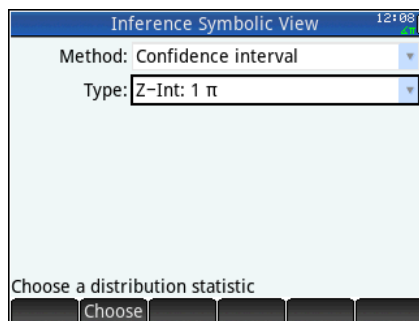
Problem 2

A calculator's random number generator is supposed to produced numbers x from a uniform distribution such that 0≤x<1. A student generates 40 random numbers from her calculator:

0.2780	0.3670	0.8001	0.7101	0.9059	0.1329	0.4847	0.0425
0.1207	0.9666	0.3123	0.6745	0.4215	0.3741	0.1997	0.3856
0.8302	0.2247	0.3871	0.1939	0.0741	0.0765	0.1973	0.6589
0.7826	0.3374	0.7739	0.7093	0.1454	0.3527	0.7528	0.2580
0.4836	0.4097	0.4691	0.7819	0.4934	0.5246	0.5438	0.3172

We know that the mean of the population should be μ=0.5., but the mean of this sample is only $\bar{x}$ =0.4489. Construct a 95% confidence interval for the mean of all random numbers generated by this calculator.

1. Get the *CIProblem2* app from your instructor. It contains the data set in D1 of Numeric view.

2. Tap ⬚Stats⬚ to see summary statistics for our sample, including $\bar{x}$ =0.4489.

We know that the sample was random. To establish that the sampling distribution is roughly normal, we look at the histogram, the box and whisker plot, and the normal probability plot.

3. Press ⬚Symb⬚ to open Symbolic view. Set **H1** to use list D1 for its data and set **Plot1** to Histogram, as shown in the figure.

4. Press ⇧Ⓢ̲ʰⁱᶠᵗ 📊 (Plot Setup) and set both **H Width** and **X Tick** to 0.1. Do not worry about the other settings. Press 📋 and select `Autoscale` instead.
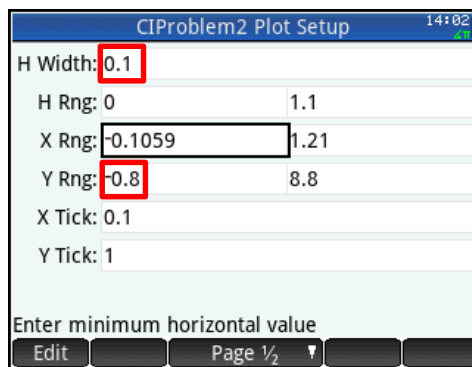
5. The histogram is displayed in Plot view. In the figure, the cursor is in the bin defined by the interval [0.3, 0.4) and the frequency of values in that range is F=8. Tap to move the tracer to any desired bin or use the directional pad.
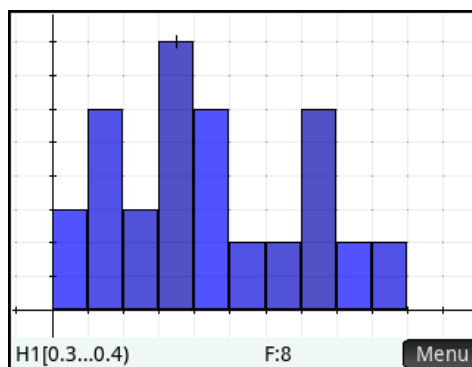
6. Let's create a box and whisker plot. Return to Symbolic view and change **Plot1** to `Box Whisker`. Under **Options**, select `Show outliers`. Again, autoscale the plot. The plot is shown in the figure to the right. Again, tap to move the tracer through the 5-number summary or use the directional pad.
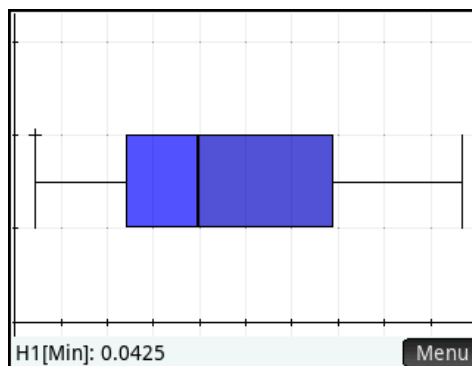
7. Finally, return to Symbolic view and set **Plot1** to `Normal Probability`. Again, autoscale the plot.

8. Do the plots show any skewness or outliers?
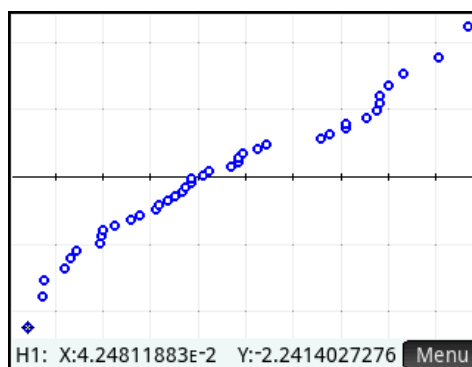
We have checked for randomness of the sample, normality of the distribution, and we know that the samples were independent. We now move to create the confidence interval.

9. Press ⊞ **Apps Info** and select `Inference`. In Symbolic view, set **Method** to `Confidence interval` and **Type** to `T-Int: 1 μ`.

10. Press ⊞ **Num↦Setup** and tap `Import`. In the **App** field, select `CIProblem2`; in the **Column** field, select `D1`. Tap `OK` to import the values shown for $\bar{x}$, s, and n. Let C=0.95. The finished Numeric view is shown in the figure to the right.

11. Tap `Calc` to see the confidence interval displayed along with the confidence level, the degrees of freedom, and the critical value of t.

12. Interpret the 95% confidence interval for the mean of all random numbers generated by this calculator. Do you think the random number generator is functioning properly?
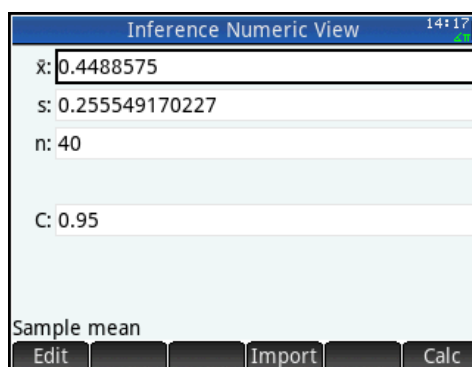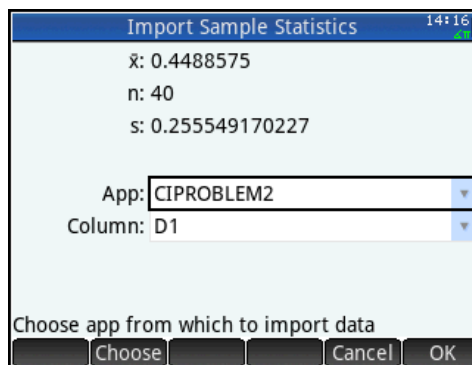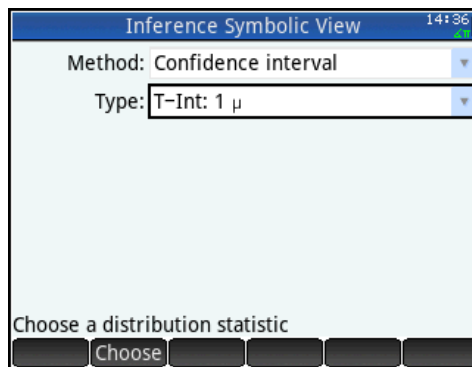
_____

_____

_____

Answers

Problem 1

1. The 95% confidence interval is (0.1942, 0.5391). We are 95% confident that the true proportion of heads for all tosses of this coin is between 0.1942 and 0.5391. Since this confidence interval contains 0.5, this coin appears to be fair.

Problem 2

9. No, none of the plots show any outliers and there is no skew present.

12. The confidence interval is (0.3671, 0.5306). We are 95% confident that the mean of all random numbers generated by this calculator is between 0.3671 and 0.5306. Since the interval contains 0.5, this calculator appears to be functioning correctly.

Teacher Notes

The confidence intervals for a difference between two means or a difference between two proportions are constructed in roughly the same manner as those for a single mean or proportion. Choose the type in Symbolic view, then go to Numeric view and either enter the data or import it from another app. Tap ⌐Calc⌐ to see the results. Examples are shown below.

Confidence interval for a difference of proportions

Confidence interval for a difference of means

The Results view is designed to give you the information you need to confirm the confidence interval calculation by hand or step-by-step. This design gives the teacher greater flexibility in the classroom.

For example, in Problem 1 we know n=30 and x=11 (number of successes). The formula for the confidence interval is

$\hat{p} \pm z^* \cdot \sqrt{\dfrac{\hat{p} \cdot (1 - \hat{p})}{n}}$. The value of $\hat{p}$ can be

calculated from x and n, and the value of z^* is given in the Results view.

Similarly, in Problem 2 we imported $\bar{x}$ =0.4489, s_x=0.2556, and n=40. The formula for the confidence interval is

$\bar{x} \pm t^* \cdot \dfrac{s_x}{\sqrt{n}}$. The value of t^* is given in the

Results view. Note that the degrees of freedom are given as well so that the value of t^* can be confirmed using STUDENT_ICDF(39,0.025).

Tests of Significance

In this activity, we look at tests of significance in a variety of contexts, using HP Prime.

HP Prime Functionality Introduced:

Using the Inference app Numeric, Symbolic, and Plot views
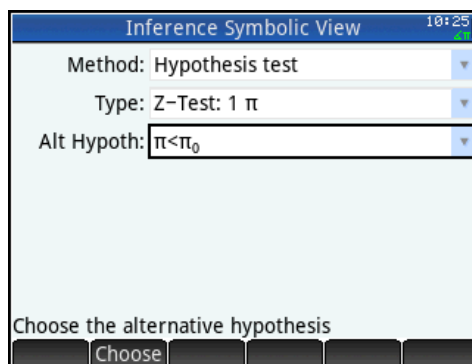
AP Statistics Content:

Large sample tests for a proportion and a difference between two proportions, test for a mean and a difference between two means

C. I. Problem 1 Revisited

Let's examine Problem 1 from the Confidence Interval activity from the standpoint of a test of significance. As you remember, a coin was tossed 30 times and came up heads 11 times. Test the null hypothesis H_0: $\pi=0.5$ against the alternative hypothesis H_a: $\pi<0.5$ at the $\alpha=0.10$ level.

1. Press **Apps Info** and select `Inference`. For **Method**, select `Hypothesis test` and for **Type**, select `Z-Test: 1 π`. For the alternative hypothesis, select $\pi<\pi_0$, as shown in the figure.

2. Press **Num Setup** to open Numeric view. From our problem, enter x=11, n=30, π_0=0.5, and α=0.1, as shown.
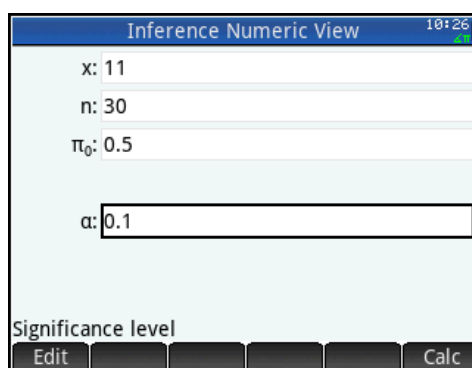
3. Tap [Calc] to see the test results.

The test result is reported as 0 or 1, to reject or fail to reject the null hypothesis. After this value, both the test z and test $\hat{p}$ values are displayed. Finally, the p-value of the test and the critical values of z and $\hat{p}$ are reported. In this case, the p-value (p=0.072) is less than the α-level (α=0.10),

so at this α-level we have evidence to support the alternative hypothesis.

4. Tap [OK] to return to Numeric view.
5. Press [Plot/Setup] to open Plot view.

Here the test result is presented graphically in two ways. The first is the default view, showing the mapping between the test z and test $\hat{p}$.

6. Tap [α] to switch to the second graphical representation of the results. A white dot will appear to show the second option is active.

Here the reject region is shaded light blue and the position of the test z-value relative to the reject region is displayed. The test value is within the reject region as expected, but it is quite close to the boundary of the region. Had we decided to use α=0.05, the test result would have been quite different.

7. Press ⊙ and ⊙ to increase and decrease the value of α, dynamically illustrating the relationship between the magnitude of α and the area of the reject region.

Results	10:29
Result	0
Test Z	-1.46059348668
Test $\hat{p}$	0.366666666667
P	0.072063517408
Crit. Z	-1.28155156554
Crit. $\hat{p}$	0.387247197167

Reject H$_0$ at α=0.1
[More] [ ] [ ] [ ] [OK]

P=0.072063517408
Test Z=-1.46059348668

0   Z
0.5   P
Test $\hat{p}$=0.366666666667
[α] [ ] [ ] [ ] [ ]

-1.28155156554 ⇐Crit. Z
Test Z=-1.46059348668

0   Z
0.5   P
Test $\hat{p}$=0.366666666667
[α•] [α=0.10] [ ] [ ] [ ]

-1.64485362695 ⇐Crit. Z
Test Z=-1.46059348668

0   Z
0.5   P
Test $\hat{p}$=0.366666666667
[α•] [α=0.05] [ ] [ ] [ ]
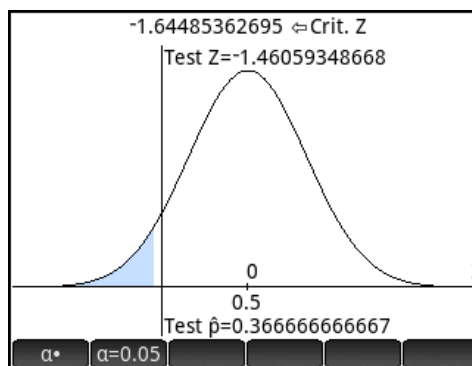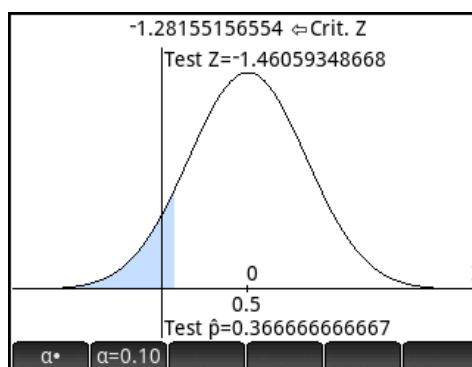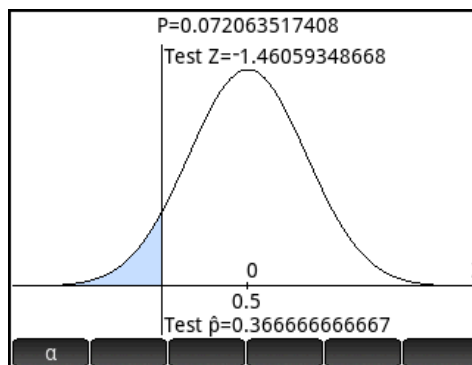
C. I. Problem 2 Revisited

A calculator's random number generator creates a sample of 40 random numbers x, with $0 \leq x < 1$. The sample has $\bar{x} = 0.4489$ and $s_x = 0.2556$. Test the null hypothesis H_0: $\mu = 0.5$ against the alternative hypothesis H_a: $\mu < 0.5$ at the $\alpha = 0.05$ level.

1. Return to Symbolic view. For **Method**, select `Hypothesis test` and for **Type**, select `T-Test: 1 μ`. For the alternative hypothesis, select $\mu < \mu_0$, as shown in the figure.
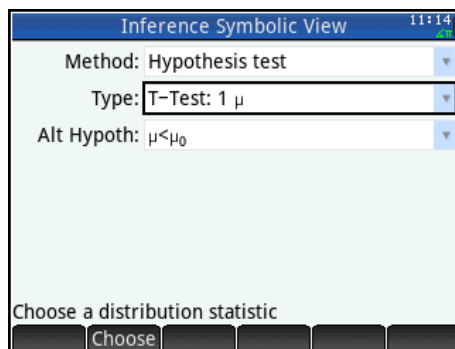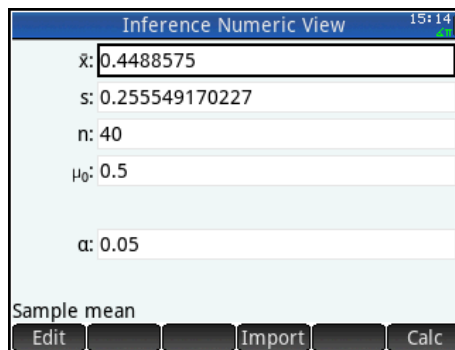
Inference Symbolic View	11:14
Method: Hypothesis test	▼
Type: T−Test: 1 μ	▼
Alt Hypoth: μ<μ₀	▼

Choose a distribution statistic
| Choose |

2. Press Num/Setup to open Numeric view. Either enter the values above or import them again from the CIProblem2 app. The finished Numeric view is shown to the right.

Inference Numeric View	15:14
x̄: 0.4488575	
s: 0.255549170227	
n: 40	
μ₀: 0.5	
α: 0.05	

Sample mean
| Edit | | | Import | | Calc |

3. Tap `Calc` to see the test results.

The test result (1) is displayed first, indicating that we should fail to reject the null hypothesis. The result is followed by the test t- and $\bar{x}$-values. These are followed in turn by the p-value of the test, the degrees of freedom, and the critical values of both t and $\bar{x}$. In this case, the p-value is greater than the α-value.

Results	15:14
Result	1
Test T	-1.2657195098
Test x̄	0.4488575
P	0.106561420621
DF	39
Crit. T	-1.68487512171
Crit. x̄	0.431921152163

Fail to reject H_0 at α=0.05
| | More | | | | OK |

Again, you can press Plot/Setup to see two graphical representations of this test result, one with the test- and p-values, and the other with the reject region shaded and the test value displayed.

-1.68487512171 ⇐ Crit. T
Test T=-1.2656377877

0
0.5
Test x̄=0.448860015687

| α• | α=0.05 | | | | |
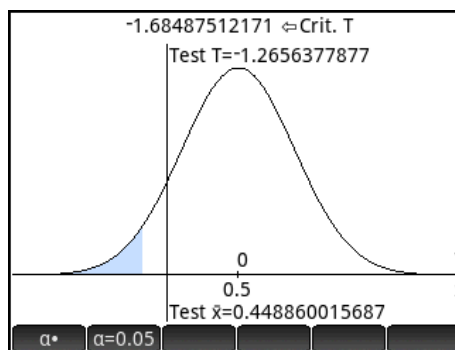
Distributions of Categorical Data

In this set of activities, we look at distributions of categorical data using the Chi-square distribution.

HP Prime Functionality Introduced:

The HP Prime Inference App Chi-square methods

AP Statistics Content:

Chi-square tests for goodness of fit and homogeneity of proportions

Activity: Comparing Observed and Expected Counts

M&M's peanut candies are packaged, on average, with the following percentages of each color:

- 23% blue
- 23% orange
- 15% green
- 15% yellow
- 12% red
- 12% brown

I opened a 300 gram bag of peanut M&M's and found 120 candies, with their colors distributed as follows:

- 22 blue
- 15 orange
- 12 green
- 36 yellow
- 20 red
- 15 brown

There seemed to be an awful lot of yellow candies! How likely is it that the distribution in my sample is due to chance variation?

1. Press **Apps Info** to open the App Library and select Inference. The app opens in Symbolic view.
2. In the **Method** field, select χ^2 test. For the **Type**, select Goodness of Fit. Note that you can select to enter either expected counts or expected probabilities. Here, we opt to enter expected probabilities.
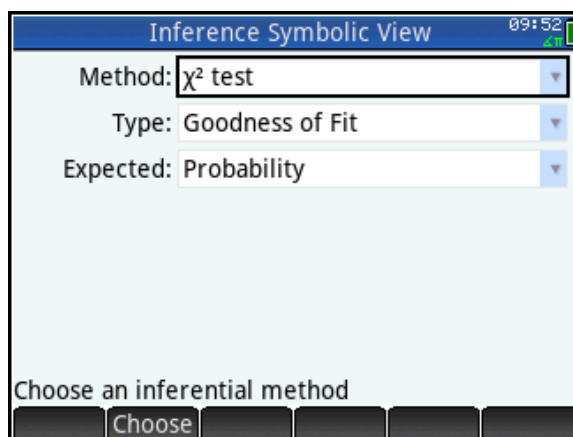
3. Press ⌘ to open Numeric view. In ObsLIst, enter the observed counts; in ProbList, enter the corresponding probabilities.

	ObsList	ProbList
1	22	0.23
2	15	0.23
3	12	0.15
4	36	0.15
5	20	0.12
6	15	0.12
7		

Enter value or expression
Edit | More | Go To | Import | Make | Calc

4. Tap Calc to perform the chi-square test of goodness of fit and view the results. On average, 1 bag in 2,500 would have a distribution at least as extreme as mine just due to chance variation.

Stats
χ^2	29.0911835749
P	2.2251780227E-5
DF	5

Calculated test probability
More | Stats• | Exp | Cont | OK

5. Tap Exp to see the expected counts. Tap Cont to see the chi-square contributions by category. As you can see in the figure to the right, much of the chi-square value was contributed by the large number of yellow candies.

6. Tap OK to return to Numeric view.

Stats
1	1.13623188406
2	5.75217391304
3	2
4	18
5	2.17777777778
6	0.025
1	

More | Stats | Exp | Cont• | OK

In general, the Inference App consistently works in the following manner:
1. In Symbolic view, choose a Method and Type
2. In Numeric view, enter or import your data
3. Tap Calc to see the results
4. Tap OK to return to Numeric view.

Activity: Two-Way Tables and the Chi-square Test for Homogeneity

The American Community Survey from the US Census recently published a brief: *The Foreign-Born Population From Africa: 2008-2012*. From the report, here are estimates of categorical data regarding the geographical origins of African immigrants in New York state vs. California.

Number (in thousands) of African-born immigrants from regions of Africa, 2008-2012						
	East Africa	Mid-Africa	North-Africa	South-Africa	West-Africa	Other
New York	17.9	4.9	35.4	6.1	85.8	13.8
California	49.3	4.5	45.3	16.0	30.8	9.1

We wish to test the null hypothesis, H_0: there is no difference in the New York and California populations with respect to geographical distribution of their African immigrants, against the alternative hypothesis, H_a: there is a difference between these two populations with respect to geographical distribution.

1. Press ⬛ to open the App Library and select Inference. The app opens in Symbolic view.

2. In the **Method** field, select X^2 test. For the **Type**, select 2-way test.

3. Press ⬛ to open Numeric view. Enter the data in ObsMat, as shown in the figure to the right.

4. Tap ⬛Calc⬛ to perform the chi-square test of homogeneity and view the results. Since our p-value is quite small, we have strong evidence to support rejecting the null hypothesis in favor of our alternative hypothesis.

5. Again, tap ⬛Exp⬛ to see the expected counts. Tap ⬛Cont⬛ to see the chi-square contributions by category.

Much of the chi-square value was contributed by the differences in the populations from East and West Africa. More West Africans seem to settle in New York, while more East Africans tend to settle in California.