

Exit Stage Left



Teacher Notes & Answers

7 8 9 10 11 **12**



TI-30XPlus
MathPrint™



Activity



Student



50 min

Teacher Notes:

There are lots of opportunities to immerse students in the theme of this activity. The activity includes two PowerPoint slide shows that contain the main data. To help engage students in the theme, try playing movie theme songs in the background as students enter the data on their calculator. Data stored on the TI-30XPlus MathPrint calculator will remain until the user decides to delete it.

A separate emulator state file: Awards.s3l included in the activity bundle contains the data from the slide show. This file can be loaded into the emulator to save teacher time at the start of the lesson.

In the TI-SmartView software select: Actions > Load Emulator State

Additional data is contained in the "Awards" Spreadsheet. The additional data is used in the extension section of this activity: Hidden Figures [2016 Movie: <http://www.hiddenfigures.com>]

The activity also includes a section titled "Miss Representation" [missrepresentation.org]. This section involves looking beyond the 40year sample from the awards data. Students use simple median smoothing and linear regression to explore if there has been any long-term change in gender bias. The median smoothing provides a contextualised platform for students to do some quick repetitive calculations, at the same time, highlighting the difference between mean and median with regards to reducing the impact of outliers.

Content covered in this activity, specific to the HSC Mathematics Standard course includes, but is not limited to:

Construct and compare parallel box-plots:

- Complete a five-number summary for different datasets (ACMEM058)
- Compare groups in terms of central tendency (median), spread (IQR and range) and outliers (using appropriate criteria)
- Interpret and communicate the differences observed between parallel box-plots in the context of the data

About the Activity

The activity starts with students recording the age of actors (male) via a series of animated Power Point slides. The TI-30XPlus MathPrint calculator is perfectly capable of calculating quartiles automatically, however it is important that students understand 'how' these are calculated. The students use the "Sort" tool in the calculator to put the ages into numerical order and then locate the minimum (Q_0), median (Q_2) and maximum (Q_4) and finally the first (Q_1) and third (Q_3) quartiles and then compare their answers to those obtained by the calculator. This is also a 'teaching moment' where teachers can discuss how Q_1 and Q_3 are calculated, whether to include or exclude the median. To emphasis the options, delete the last data value, now the list will only contain 39 entries. This will make the median easy to determine, but students will need to consider whether to include or exclude it in their calculations for Q_1 and Q_3 . With only 39 data points, students can check to see whether the calculator includes or excludes the median. Students then determine the Inter-quartile range (IQR) to see if their data contains any outliers, again, very easy given the data has been sorted. From here, students draw a boxplot on the provided grid. The data for female awards is then presented, recorded, analysed and statistically compared to male ages. The significant contrast in the two boxplots is incredible!

Introduction

The term “exit stage left” refers to an uneventful departure, making way for more interesting events. If you’re female and working in the Hollywood, you might be mistaken for thinking the definition said “youthful” rather than “interesting”. When a 28 year old actress is knocked back because she is too old to play the love interest of a 39 year old actor, and a 37 year old actress too old for the 55 year old actor, something is clearly wrong. Could all of this be mumurs and rumbings from a small gendered group of disgruntled celebrities or is there data to support these claims?

In 2004 Geena Davis founded the “Geena Davis Institute on Gender in Media”, the organisation has been busy collecting data. Now it’s your turn, is this a case of “Entrapment” (1999) or “Somethings Gotta Give” (2003). The data source used for this investigation comes from the pinnacle of the movie industry: the Academy Awards®. These awards commenced in May 1928. Since 1935, Pricewaterhouse has managed the awards and initiated the secrecy of the result by placing the winner’s name in a sealed envelope, coining the phrase: “The envelope please”. Past award winners include Jodie Foster (29), Gwyneth Paltrow (26), Julia Roberts (33), Will Smith (52), Jack Nicholson (60) and Anthony Hopkins (82).

Data Collection: Males

This activity uses all three of the TI-30XPlus MathPrint lists; it is therefore appropriate to clear all the lists before starting.

Press:  ... select option 4, **Clear All**.

The first set of data to be explored is the age at which ‘actors’ (male) received their award. This activity uses the past 40 years as a sample from almost 100 years of awards.

The data for the male ages will be stored in List 2.

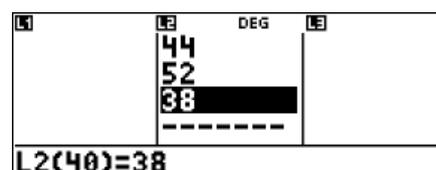
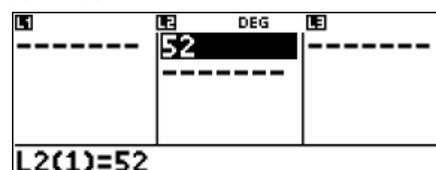
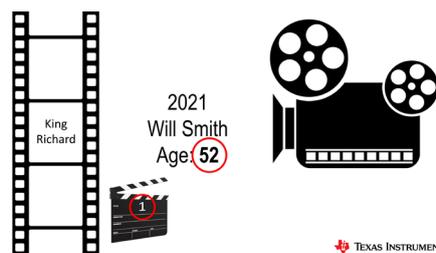
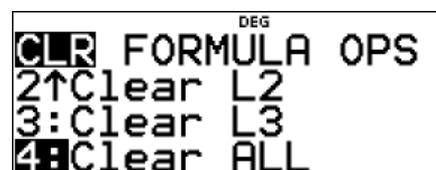
Navigate to list 2. Your teacher will present the data using a series of PowerPoint slides. Each slide contains the year, actor’s age (at the time of the award) and an item number. The item number should match the list item number on your calculator.

In the calculator screen image (opposite), Will Smith’s age has been entered as the first item in list 2: $L2(1) = 52$

Match the item number on each slide to make sure your data is correct.

Data Entry

Once you have finished entering the data, your screen should show the last item as: $L2(40) = 38$.



Median

The median is the middle number in a sorted list, this means that 50% of the data lies above the median and 50% below. If the quantity of items in the list is even, the ‘average’ of the middle two items is calculated.

Sorting the List

We want to retain the current order for the data so the sorted data will be sent to a separate 'working' list.

stat-reg/distr

Press: **data** **←** ... select option 1: Sort Smallest to Largest.

Our data is located in List 2 (L2).

The sorted data will be sent to List 1 (L1).

Once these selections have been made, select "SORT".

An ordered copy of List 2 will be sent to List 1.

We can see from the sorted list that the youngest actor to win an award in this forty-year period is 28 years of age. Navigate down to the 20th item in List 1.

Remember: There is an even quantity of data in the list.

```

DEG
CLR FORMULA OPS
1:Sort Sm-Lg...
2:Sort Lg-Sm...
3↓Sequence...
  
```

```

DEG
SORT SMALL-LARGE ↑
SORT LIST: L1 L2 L3
→ LIST: L1 L2 L3
SORT
  
```

L1	L2	DEG	L3
28	52		-----
31	82		
31	44		
32	36		

L1(1)=28

Question: 1.

Use your sorted list to help determine the following statistics for male award winners over the past 40 years.

- Minimum Age (Q_0) **Answer: 28** (First list item in L1)
- Maximum Age (Q_4) **Answer: 82** (Last list item in L1) ... Tip: Press 2nd + Down arrow.
- Median Age (Q_2) **Answer: Item 20 = 43, Item 21 = 44, therefore the median = 43.5**
- First Quartile (Q_1) **Answer: Item 10 = 37, Item 11 = 37, therefore $Q_1 = 37$**
- Third Quartile (Q_3) **Answer: Item 30 = 50, Item 31 = 51, therefore $Q_3 = 50.5$**

The calculator can determine each of these quantities using the statistics menu.

stat-reg/distr

Press: **2nd** **data** ... select option 2 (1 Variable Statistics)

The original data is located in List 2 (L2). The data does not include any frequency information, therefore, select "ONE" for the Frequency option, the select **calculate** (CALC).

```

DEG
STAT-REG DISTR
1:StatVars
2:1-VAR STATS
3↓2-VAR STATS
  
```

```

DEG
1-VAR STATS ↑
DATA: L1 L2 L3
FREQ: ONE L1 L2 L3
CALC
  
```

Question: 2.

The mean age for male award winners over the past 40 years is approximately 44.5. Suggest reasons why this statistic is greater than the median for the same set of data.

Answer: The data may contain outliers, it may be skewed, or both. The median is less likely to be impacted by skewed data.



Outliers

An outlier for a boxplot is a data point that lies beyond the whiskers. To calculate if a data point is an outlier, determine the Inter Quartile Range ($IQR = Q_3 - Q_1$) and then calculate the lower and upper fence.

Lower fence: $Q_1 - 1.5 \times IQR$

Upper fence: $Q_3 + 1.5 \times IQR$

Question: 3.

Use your previous calculations to help determine each of the following:

- a) Inter-Quartile Range (IQR). **Answer:** $IQR = 50.5 - 37 = 13.5$
 b) Lower Fence. **Answer:** 16.75
 c) Upper Fence. **Answer:** 70.75
 d) Which data points (if any) are outliers? **Answer:** 82

**Tip!**

stat-reg/distr A
 Press: **2nd** **data** **1**

Navigate to Q₃ and Q₁ to paste these values to the calculator.

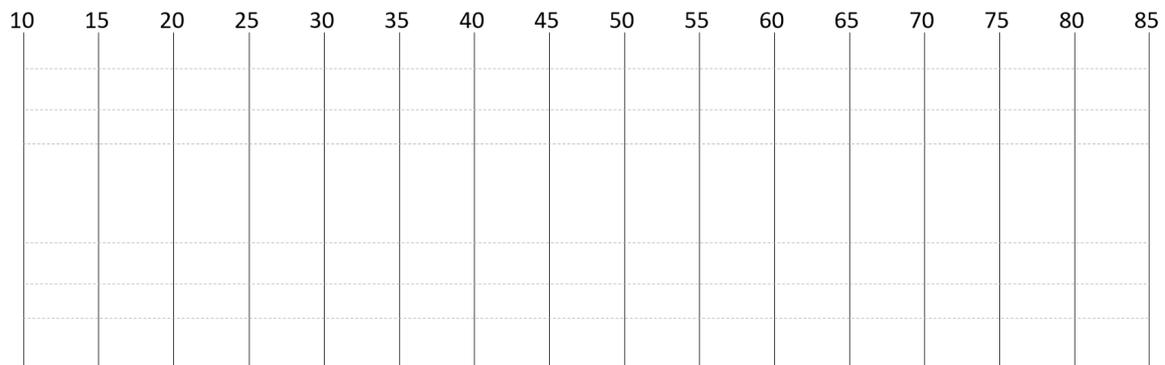
Q3-Q1

DEG

Question: 4.

Use the statistics determined previously, to draw a boxplot for the male award winners over the past 40 years.

Note: The additional space is for the female data to be added later.

**Data Collection: Females**

It's time to enter the data for the actresses. The slide set is in the same format. Use the clapper-board to make sure your data is correctly aligned.

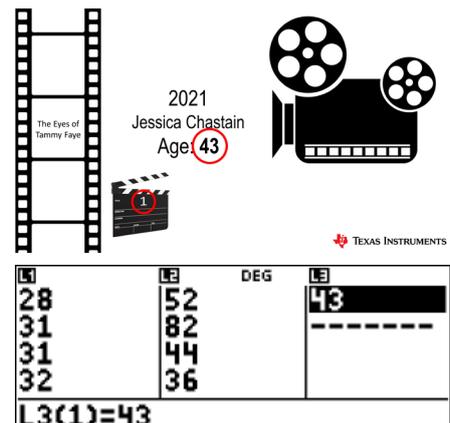
Store your data in List 3 (L3).

Do **NOT** delete the data for males, this will be used again later!

stat-reg/distr

Press: **data** ... navigate across to List 3 and enter the data.

The calculator screen opposite shows the first entry: 43 that has been entered into list 3: $L3(1) = 43$



Question: 5.

Use the statistics functionality of the calculator to determine the following statistics for female award winners over the past 40 years.

- a) Minimum Age (Q_0) **Answer: 20**
- b) Maximum Age (Q_4) **Answer: 79**
- c) Median Age (Q_2) **Answer: 34**
- d) First Quartile (Q_1) **Answer: 28.5**
- e) Third Quartile (Q_3) **Answer: 46**

Question: 6.

The mean age for female award winners over the past 40 years is approximately 38.9, just over five and a half years younger than their male counter-parts. Discuss whether you believe this to be a significant difference.

Answer: Answers will vary. Students should reference the median in their discussion. There is almost a 10-year difference in the median values! Quartiles may also be used to support a specific point of view.

Teacher Notes: The purpose of this question is to illustrate that sometimes a single statistic is insufficient to provide details pertaining to the 'big picture'. Students may change their mind a little once they complete the boxplot.

Question: 7.

Use your previous calculations to help determine each of the following for females:

- a) Inter-Quartile Range (IQR). **Answer: $IQR = 46 - 28.5 = 17.5$**
- b) Lower Fence. **Answer: 11**
- c) Upper Fence. **Answer: 63.5**
- d) Which data points (if any) are outliers? **Answer: 79**

Question: 8.

Use the statistics determined to draw a boxplot for female award winners over the past 40 years.

Note: Use the additional space provided adjacent to the male boxplot.

Question: 9.

How do the adjacent boxplots support the concerns expressed by the Geena Davis Institute on Gender in Media?

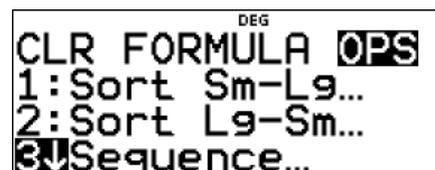
Answer: The parallel boxplots provide a visual comparison making it much clearer to see that Q_1 for males lies clearly above the median for females. Statistically we can say that more than 50% of female award winners (in our sample) are below the age of 37 (Q_1) for males.

Miss Representation

Our boxplots may be misrepresenting the data as they don't show trends over time.

To explore if any changes have occurred, the years need to be added, these will be stored in List 1 (L1).

Press:    and select option 3 (Sequence)



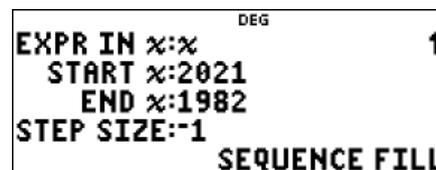
Select List 1 (L1) for the time series information.



Match the settings shown opposite.

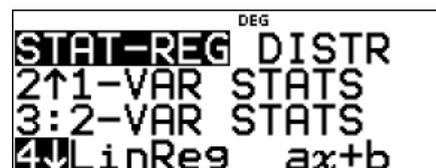
The first data entry is for 2021, the last one is for 1982, so our dates are stepping backwards (-1).

Select SEQUENCE FILL and the years will be sent to List 1.



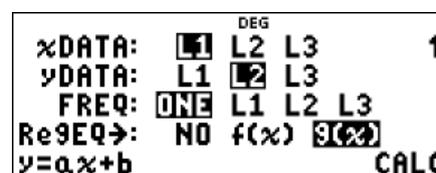
It's now time to see if there are any trends in our data.

Press: 2nd stat-reg/distr data ... select option 4: LinReg $ax + b$



The "years" data is located in List 1 (L1), the ages for males located in List 2 (L2), store the regression equation in $g(x)$

Match the settings shown opposite then select **CALC**.



Question: 10.

Determine the regression equation and comment on whether there is any trend in the data for the males.

Answer: Equation: $y = 0.164x - 285$. Pearson's correlation coefficient: $r^2 = 0.035$, indicates there is effectively no correlation in the data. ie: No trend.

Question: 11.

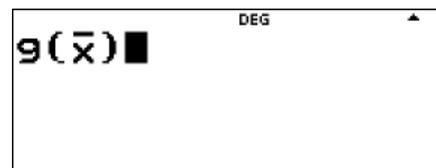
The regression equation has been stored in $g(x)$.

- a) Calculate $g(\bar{x})$ using the Table tool and Stat-Vars where \bar{x} represents the 'average' (mean) year.

Answer: $g(\bar{x}) = 44.475$

- b) Recall and compare the average (mean) age with the previous calculation.

Answer: $\bar{y} = 44.475$ The least squares regression line passes through (\bar{x}, \bar{y}) .



Question: 12.

Determine the regression equation for the females and comment on whether there is any trend.

Answer: Equation: $y = 0.123x - 207$. Pearson's correlation coefficient: $r^2 = 0.011$, indicates there is effectively no correlation in the data. ie: No trend.

Question: 13.

The regression equation has been stored in $f(x)$, calculate and compare $f(\bar{x})$ and compare the result to the average age of female award winners.

Answer: $f(\bar{x}) = 38.9$

Hidden Figures

The data investigated in this activity has focused only on the past 40 years. Trends may have occurred over longer periods of time, including movie directors and scripts incorporating leading roles for older females. For example, in the 1980's Katherine Hepburn (74), Shirley MacLaine (49), Geraldine Page (60) and Jessica Tandy (79) all won awards.

To analyse this data, "5 point median smoothing" will be applied.

Example:

Year:	1928	1929	1930	1931	1932
Winner:	Janet Gaynor	Mark Pickford	Norma Shearer	Marie Dressler	Helen Hayes
Age:	21	37	27	62	31

Question: 14.

Using the above table:

a) Calculate the mean. **Answer: Mean = 35.6**

b) Calculate the median **Answer: Median = 31**

c) Which statistic, mean or median best represents the data?

Answer: Answers will vary, however the 62 appears to be an outlier, even amongst this small group. Amongst the other four winners there is only a 16-year age range. Marie was 25 years older than the next oldest in this group. The 'median' calculation effectively ignores this value as an outlier.

Question: 15.

Collect the median data for the life span of the awards for both males and females and determine the following:

a) The regression equation for the 5 point median data for females.

Answer: Equation for females: $y = 0.098x - 160$. $r^2 = 0.34$

b) The regression equation for the 5 point median data for males.

Answer: Equation for males: $y = 0.037x - 29$. $r^2 = 0.052$

c) Discuss the two equations and the corresponding correlation.

Answer: The data for the males still appears to show no correlation ($r^2 = 0.052$), there is however a weak correlation for the female data. The 'gradient' for the female model is approximately 0.1, the equation predicts a small increase in the age for award winners, 1 year in 10, however given the correlation strength, even this marginal increase is questionable.

Teacher Notes: A median-median regression line might also be appropriate here given that the "Geena Davis Institute on Gender in Media" launched in 2004. Has it had an impact? Our 40 years of data straddles the launch date quite nicely. Removing one year of data leaves 39 points, so each section of the median-median calculations will consist of 13 data points. Alternatively, shrink the data set so that 2004 lies in the middle of the data.