

# The IB HL or MM Statistics Option with the TI-83

Version 2.6 - corrections & additions welcome - Dr. Wm J. Larson - [william.larson@ecolint.ch](mailto:william.larson@ecolint.ch)

## Table of Contents

<b>Creating a list</b> .....	<b>1</b>
<b>Calculating Mean, Median, Range, Standard Deviation &amp; Interquartile Range</b> .....	<b>1</b>
Mean & Median	1
Range	1
Standard Deviation	1
Interquartile Range	1
Redisplaying Data	1
Be careful to clear the screen	1
<b>Calculating Probabilities for the Normal Distribution</b> .....	<b>1</b>
Using ShadeNorm	1
Using normalcdf	2
Significant Digits	2
<b>Calculating the Inverse Normal Distribution</b> .....	<b>2</b>
Using invNorm	2
<b>Calculating Probabilities for the Poisson Distribution (Higher Level only)</b> .....	<b>2</b>
Using poissonpdf	2
<b>Confidence Intervals</b> .....	<b>2</b>
Calculating a Z interval	2
<b>Hypothesis Testing</b> .....	<b>3</b>
Conducting a Z-Test	3
Conducting a t-Test (Higher Level only)	3
Conducting a $\chi^2$ Test for Independence i.e. Contingency Tables	3
Conducting a $\chi^2$ Test for Independence with the Yates Continuity Correction	3
Conducting a $\chi^2$ Goodness of Fit Test (Higher Level only)	3
<b>Regression and Correlation Analysis</b> .....	<b>3</b>
Drawing a Scatter Diagram	3
Fitting a line	3
Covariance	4
The equations you can fit:	4

## Creating a list

Key STAT, EDIT & just type your data in. Or in the home screen you can type, for example, {1,1,2,2,3,4} STO⇒ L1

## Calculating Mean, Median, Range, Standard Deviation & Interquartile Range

### Mean & Median

First enter your data into a list, say L1. See above.

To get the mean key 2<sup>nd</sup> List, Math 3: mean(L1), Enter

If the frequency is not 1 and is in L2, key mean(L1, L2), Enter.

To get the median key 2<sup>nd</sup> List, Math 4: median(L1), Enter

If the frequency is not 1 and is in L2, key median(L1, L2), Enter.

**Or** key STAT, CALC, 1: 1-Var Stats, L1, Enter. A set of data about L1 called 1-Var Stats will appear. The mean,  $\bar{x}$ , is at the top of the list. Scrolling down the median, Med, will be displayed. L1 is the default, so if the data is in L1, you need not type "L1".

### Range

First enter your data into a list, say L1. See above.

Range = Maximum - Minimum

To get the maximum key 2<sup>nd</sup> List, Math 2: max(L1), Enter.

To get the minimum key 2<sup>nd</sup> List, Math 1: min(L1), Enter.

**Or** key STAT, CALC, 1: 1-Var Stats, L1, Enter. A set of data about L1 called 1-Var Stats will appear. Scrolling down minX, & maxX will be displayed.

### Standard Deviation

First enter your data into a list, say L1. See above.

To get the minimum key 2<sup>nd</sup> List, Math 1: stDev(L1), Enter. stDev gives the sample standard deviation, which is not used in math methods.

**Or** Key STAT, CALC, 1: 1-Var Stats, L1, Enter. A set of data about L1 called 1-Var Stats will appear, including Sx, the sample standard deviation, &  $\sigma_x$ , the population standard deviation. Only the population standard deviation is used in math methods. Both the sample standard deviation and the population standard deviation are used in higher level.

### Interquartile Range

First enter your data into a list, say L1. See above.

Interquartile range = Q3 – Q1.

The 1<sup>st</sup> quartile = Q1 is the median of the lower ½ of the data, including the median.

The 3<sup>rd</sup> quartile = Q3 is the median of the upper ½ of the data, including the median.

To get the interquartile range key STAT, CALC, 1: 1-Var Stats, L1, Enter. A set of data about L1 called 1-Var Stats will appear. Scrolling down Q1 & Q3 will be displayed.

### Using a frequency list

If you are given data points with frequencies for each data point, put the data points in L1 & the frequencies in L2. Then key STAT, CALC, 1: 1-Var Stats, L1, L2.

L1 is the default for the data list, so if there is no frequency list & the data is in L1, you need not type "L1". But there is no default for the frequency list. So if there is a frequency list in L2, you need to type 1-Var Stats L1, L2.

### Redisplaying Data

If you cleared the screen (but did not run a new statistics calculation), you can redisplay your data. For example you can redisplay Q1 & Q3 by keying VARS 5:Statistics, PTS & then selecting 7:Q1 or 9:Q3.

### Be careful to clear the screen

The TI-83 has a tendency to display information from a previous calculation, so when you are making a new calculation, always clear the screen first using CLEAR, CLEAR.

## Calculating Probabilities for the Normal Distribution

### Using ShadeNorm

ShadeNorm will draw the graph and calculate the probability.

Key 2<sup>nd</sup> DISTR DRAW1: ShadeNorm(lowerbound, upperbound [,  $\mu$ ,  $\sigma$ ])

### Example

Find  $P(z < -0.5)$ . (The default vales of  $\mu = 0$ ,  $\sigma = 1$  are desired, so they need not be entered.)

Key DISTR DRAW 1: ShadeNorm(-100, -.5)

The graph, the lower bound (-100, being 100 standard deviations from the mean, is effectively minus  $\infty$ ), the upper bound and the  $P(z < -0.5)$ , i.e. 0.3085 are displayed.

If the graph is not visible, set the Window to:

```
xmin = -3
Xmax = 3
Xscl = 1
Ymin = -.25
Ymax = .5
Yscl = .25
Xres = 1
```

**Example**

If  $\mu = 55$ ,  $\sigma = 10$ , find  $P(x < 65)$ .

Key ShadeNorm(-E99, 65, 55, 10).

The graph, the lower bound ( $-1 \times 10^{99}$ ) & the upper bound,  $P(x < 65)$ , i.e. 0.841 are displayed.

Unless you first keyed 2<sup>nd</sup> DRAW 1: ClrDraw, the graph may not be redrawn from a previous graph, although the numbers on the bottom will be correct.)

If the graph is not visible, set the Window to:

```
xmin = 25 (i.e.  $\mu - 3\sigma$ )
Xmax = 85 (i.e.  $\mu + 3\sigma$ )
Xscl = 10 (i.e.  $\sigma$ )
Ymin = -.025 (i.e.  $-1/4\sigma$ )
Ymax = .05 (i.e.  $1/2\sigma$ )
Yscl = .025 (i.e.  $1/4\sigma$ )
Xres = 1
```

**Example**

If  $\mu = 55$ ,  $\sigma = 10$ , find  $P(40 < x < 65)$ .

Key ShadeNorm(40, 65, 55, 10).

**Using normalcdf**

normalcdf will only calculate the probability, i.e. the **normal cumulative probability distribution function**.

Key 2<sup>nd</sup> DISTR DISTR (the default)

2: normalcdf(lowerbound, upperbound [,  $\mu$ ,  $\sigma$ ])

**Example**

If  $\mu = 55$ ,  $\sigma = 10$ , find  $P(x < 65)$ .

Keying DISTR DISTR normalcdf (-E99, 65, 55, 10) will display 0.841.

We usually want the cumulative distribution function (cdf) for the normal distribution. The **probability distribution function (pdf)** would be useful to graph the normal curve in  $Y=$ , but ShadeNorm already does that.

**Significant Digits**

Notice that more significant digits are available with the TI-83 than with a normal distribution table in a textbook. However in the real world  $\mu$  &  $\sigma$  are usually not known with enough accuracy to make this meaningful.

**Calculating the Inverse Normal Distribution**

**Using invNorm**

For  $\Phi(a) \equiv P(Z < a)$  if  $\Phi$  is known but a is not known, invNorm will calculate a.

Key 2<sup>nd</sup> DISTR DISTR (the default)

3: invNorm(area, [,  $\mu$ ,  $\sigma$ ])

**Example**

If  $P(Z < a) = .6$ , find a. (The default vales of  $\mu = 0$ ,  $\sigma = 1$  are desired, so they need not be entered.)

Keying DISTR 3: invNorm(.6) will give 0.253347

**Example**

If  $x \sim N(100, 5^2)$  &  $P(x < a) = .20$ , find a.

Keying DISTR 3: invNorm(.2, 100, 5) will give 95.8.

**Calculating Probabilities for the Poisson Distribution (Higher Level only)**

**Using poissonpdf**

Since the Poisson distribution is discrete, either the cumulative distribution function (cdf) or the probability distribution function (pdf) would be useful. Use the pdf to find the probability that one value is observed ( $X = X_0$ ) & the cdf to find the probability that one of a range of values is observed ( $X \leq X_0$ ).

Key 2<sup>nd</sup> DISTR DISTR (the default), B: poissonpdf( $\mu, x$ ) or

Key 2<sup>nd</sup> DISTR DISTR (the default), C: poissoncdf( $\mu, x$ )

**Example**

If  $\mu = 3.75$ , find  $P(x = 6)$ .

Keying DISTR DISTR poissonpdf (3.75,6) will display 0.0908.

**Example**

If  $\mu = 1.4$ , find  $P(x \geq 2) = 1 - P(x \leq 1)$

Keying DISTR DISTR poissoncdf (1.4,1) will display 0.408.  $P(x \geq 2) = 1 - 0.408 = 0.592$ .

**Confidence Intervals**

**Calculating a Z interval**

Zinterval can be used to calculate a Confidence Interval. You can enter your entire sample & have the TI-83 calculate  $\square$  or you can enter  $\square$  directly.

Key STAT TESTS 7: Zinterval. Then if you are entering  $\square$  directly select Stats & key ENTER. Then enter  $\sigma$ ,  $\square$ , n & the desired confidence level (as a decimal, not as a % - it's called the "C-Level"), select Calculate & key ENTER.

If you are given the actual sample numbers, i.e. not  $\square$ , enter them into a list and then you can either calculate  $\square$  as described above (key STAT, CALC, 1: 1-Var Stats, L1) & then use Zinterval Stats.

Or you can use Zinterval Data. In Data you must enter  $\sigma$ , n & the desired confidence level as before, but instead of  $\square$  you enter the name of the list containing your data, e.g. L1, select Calculate & key ENTER.

## Hypothesis Testing

### Conducting a Z-Test

Z-Test is used to test a hypothesis. You can enter your entire sample & have the TI-83 calculate  $\bar{x}$  or you can enter  $\bar{x}$  directly. Key STAT, select TEST 1: Z-Test. Then if you are entering  $\bar{x}$  directly, select Stats & key ENTER. Then enter  $\mu_0$ ,  $\sigma$ ,  $\bar{x}$ , n & the alternative hypothesis. Select Calculate & key ENTER.

If you are using the actual sample numbers, i.e. not  $\bar{x}$ , enter the data into a list and then use Z-Test Data. In Data you must enter  $\mu_0$ ,  $\sigma$ , n & the alternative hypothesis as before, but instead of  $\bar{x}$  you enter the name of the list containing your data, e.g. L1, select Calculate & key ENTER.

### Conducting a t-Test (Higher Level only)

t-Test is used to test a hypothesis. It is more realistic than the z test in that s, the standard deviation calculated from the sample, is used, but it requires that the sample be approximately normal or large. For large samples the z & t tests give the same answer.

You can enter your entire sample & have the TI-83 calculate  $\bar{x}$  & s or you can enter  $\bar{x}$  & s directly. Key STAT, select TEST 2: t-Test. Then if you are entering  $\bar{x}$  & s directly, select Stats & key ENTER. Then enter  $\mu_0$ ,  $\bar{x}$ , Sx (i.e. s) n & the alternative hypothesis. Select Calculate & key ENTER.

If you are using the actual sample numbers, i.e. not  $\bar{x}$  & s, enter the data into a list and then use Z-Test Data. In Data you must enter  $\mu_0$ ,  $\sigma$ , n & the alternative hypothesis as before, but instead of  $\bar{x}$  you enter the name of the list containing your data, e.g. L1, select Calculate & key ENTER.

### Conducting a $\chi^2$ Test for Independence i.e. Contingency Tables

$\chi^2$ -Test is used to test a hypothesis of independence with a 2-way contingency table.

First enter your data in a matrix. Key MATRIX, select EDIT, select a matrix to fill or edit, key ENTER, change the r x c (number of rows & columns), if necessary & enter your data.

Now key STAT, TESTS, C:  $\chi^2$ -Test. Key in the name of the matrix containing your data (Observed) and the name of the matrix where you want the expected values placed by keying MATRIX NAMES, selecting the desired matrix name and keying ENTER. Otherwise use matrices A & B which will appear by default as the observed & expected matrices.

Then choose how to display your results: Draw or Calculate. Draw will draw the  $\chi^2$  distribution, and report  $\chi^2$  (the value of  $\chi^2$ ) & P (the probability of the observed values, if the null hypothesis of independence were true). Calculate will report  $\chi^2$ , P & df (the number of degrees of freedom). To view the expected value matrix, key MATRIX, EDIT 2:B (assuming you used B, the default). Note that for a  $\chi^2$  test  $df = (r - 1)(c - 1)$ .

### Conducting a $\chi^2$ Test for Independence with the Yates Continuity Correction

When the  $df = 1$ , i.e. when the observed is a  $2 \times 2$  table, the IB requires that Yates Continuity Correction be applied.  $\chi^2$

$$(\text{corrected}) \equiv \sum \frac{(|Obs - Exp| - 0.5)^2}{Exp}$$

observed data in a matrix, say [A]. Make sure matrix B is set

to be  $2 \times 2$ , using Matrix EDIT and keying in  $2 \times 2$ . Key STAT, TESTS, C:  $\chi^2$ -Test [A] [B] ENTER. Unfortunately I have not been able to find a way to get the TI-83 to do the Yates Continuity Correction, so now you have to do it by hand. Copy out the 4 expected values from [B] & do the

$$\text{math. Example suppose the Observed is } A = \begin{bmatrix} 18 & 10 \\ 8 & 14 \end{bmatrix}$$

$$\text{The TI-83 will give the expected } B = \begin{bmatrix} 14.56 & 13.44 \\ 11.44 & 10.56 \end{bmatrix} \text{ . So}$$

now by hand do  $(|18 - 14.56| - 0.5)^2/14.56 + (|10 - 13.44| - 0.5)^2/13.44 + (|8 - 11.44 - 0.5|)^2/11.44 + (|14 - 10.56| - 0.5)^2/10.56$ . Luckily it turns out that the numerator of these 4 terms is always the same for a  $2 \times 2$  table, in our example 8.6436. So you only need to calculate  $8.6436 \times (1/14.56 + 1/13.44 + 1/11.44 + 1/10.56) = 2.81$ . Now go to the  $\chi^2$  table & find that for  $df = 1$  the critical 5% value is 3.841. Since  $2.81 < 3.841$ , we fail to reject (accept) the assumption of independence.

### Conducting a $\chi^2$ Goodness of Fit Test (Higher Level only)

A Goodness of Fit Test tests whether the population fits a model, e.g. binomial, Poisson, uniform, normal, etc. The normal, binomial, Poisson, & geometric probability distributions are in 2<sup>nd</sup> DISTR. There is no  $\chi^2$  Goodness of Fit function in the TI-83, but it is easy to calculate. Put the Observed Values in L1 and the Expected Values (the values that you would get if the model you are testing is correct) in L2. In L3, enter the formula  $(L1 - L2)^2 / L2$ . (To enter a formula scroll up to L3, key ENTER & type it in.)

To find the  $\chi^2$  test statistic, enter  $\text{sum}(L3)$ . To find p, enter  $\chi^2\text{cdf}(\text{sum}(L3), E99, df)$ .  $\chi^2\text{cdf}$  is in 2<sup>nd</sup> DISTR. E99 is a very good approximation to  $\infty$ . Note that for a best fit model  $df = k - m - 1$ , where  $k =$  the number of data categories and  $m =$  the number of parameter values estimated on the basis of the sample data.

## Regression and Correlation Analysis

### Drawing a Scatter Diagram

First enter your data into lists. See above. Then Key 2<sup>nd</sup> STAT PLOT, choose a Plot, ENTER, Select ON, Type: scatter (the squiggle of dots in the upper left), the names of your x & y lists (E.g. L1 & L2 - note that these are 2<sup>nd</sup> 1 & 2<sup>nd</sup> 2). Then Key GRAPH and ZOOM 9: ZoomStat.

### Fitting a line

Eleven kinds of regressions for fitting data to a particular type of equation are available. Only 8: LinReg(a+bx) is needed for the IB. Each of them except D accept the following optional parameters Xlistname, Ylistname, freqlist, regeq. regeq is where the fitted regression equation will be stored. The defaults are L1, L2, 1, RegEQ.

If you type the independent variable into L1 & the dependent variable into L2, you can use the defaults, i.e. avoid keying in the list names. It is useful to have the regression equation, so that you can plot it on top of the scatterplot to see if the fit looks good.

You can paste regeq to Y1 by going to Y1 in Y= and then keying VARS 5: Statistics EQ 1: RegEq. Or if you want the Equation saved to Y1 instead of RegEQ, key LinReg(a+bx) L1, L2, Y1. (Or whatever lists & equation you are using.) Note if you are using the defaults (L1, L2, freqlist = 1) they

are not needed. The commas between L1, L2 & Y1 are required. "Y1" must be keyed as VARS Y-VARS 1: Function Y1. For example key LinReg(ax+b) Y1 (Y1 is in the VARS, Y-VARS, 1:Function menu.)

To get  $r$  &  $r^2$  to appear in the screen, set the diagnostics on by keying 2<sup>nd</sup> CATALOG, (x<sup>-1</sup> - to get to d faster), DiagnosticOn, ENTER, ENTER.

Key STAT, CALC, 8: LinReg(a+bx), ENTER. a, b,  $r^2$  & r are displayed.

### **Covariance**

Covariance =  $\Sigma xy - \bar{x} \bar{y}$ . Covariance can be calculated from the data displayed by STAT CALC 2: 2-VAR STATS L1, L2. Scrolling down will display  $\Sigma xy$ ,  $\bar{x}$  &  $\bar{y}$ .

### **The equations you can fit:**

3: MedMed (a sophisticated linear regression which is less sensitive to outliers than LinReg)

4: LinReg(ax+b) (the standard linear regression)

5: QuadReg (a quadratic regression  $\{y = ax^2 + bx + c\}$ )

6: CubicReg (a cubic regression  $\{y = ax^3 + bx^2 + cx + d\}$ )

7: QuartReg (a quartic regression  $\{y = ax^4 + bx^3 + cx^2 + dx + e\}$ )

8: LinReg(a+bx) (a duplication of 4, but useful because some textbooks use one definition of a linear equation, some the other. **The IB uses this one.**)

9: LnReg (a logarithmic regression  $\{y = a + b \ln x\}$ )

0: ExpReg (an exponential regression  $\{y = a b^x\}$ )

A: PwrReg (a power law regression  $\{y = a x^b\}$ )

B: Logistic (a logistic regression  $\{y = c/[1 + a e^{-bx}]\}$ )

C: SinReg (a sinusoidal regression  $\{y = a \sin(bx + c) + d\}$ )

Once the data are keyed in, trying several different regressions (not needed for IB) is very quick and easy. The one with the  $R^2$  closest to 1 is the best fit to the data.

(Equations with a number of parameters - a, b, c, etc. - equal to or greater than the number of data points should give  $r^2 = 1$ , but are not normally considered as valid candidate equations.)

The residuals (not needed for IB) are stored in the list RESID & in EQ 1: RegEQ.