Chapter 9

Using statistics

Contents

Using statistics

Introduction

This chapter describes how to use the HP 49G to analyze data statistically. You can use the HP 49G to analyze two broad categories of statistics:

- **Descriptive statistics** enables you to calculate values such as the mean, the variance, and the standard deviation. You can also apply regression techniques to the data to fit a symbolic model to it.
- **Inferential statistics** enables you to calculate values such as confidence intervals. You can also perform hypothesis tests based on the Normal Z- and Student's t-distributions.

  The inferential statistics applications includes online help. On any of the inferential statistics screens, press HELP to display help on how to complete the screen.

Descriptive statistics

Use the descriptive statistics applications of the HP 49G to analyze data stored in a matrix.

- Use the **Single-variable statistics** application to calculate values such as the mean, the standard deviation, and the variance for a set of single-variable statistics, for example one column of a matrix.
- Use the **Frequencies** application to determine the frequency distribution for a set of data.
- Use the **Fit data** application to quantify the relationship between the data in two columns.
- Use the **Summary statistics** application to calculate summaries that relate to a set of bivariate data.

Starting an application and specifying the data

To start a descriptive statistics application:

1. Press ⮣ STAT to display the Statistics choose list.
2. Use the arrow keys to select the statistics application that you want, and press ENTER.

   The input form for the application is displayed.

When you open a statistics application input form, the default data displayed is data that is contained in the ΣDAT variable. From the statistics application input forms, you can do either of the following to specify data to analyze.

- To create new data to analyze, press EDIT to open Matrix Writer. The data you create is stored in the ΣDAT variable.
- To select an object, for example an existing matrix, press CHOOS, and select the object from the list. The data is copied to the ΣDAT variable.

Single-variable statistics

You specify the column of data within the matrix to analyze.

To calculate single-variable statistics, use the following procedure:

1. Use the method described in “Starting an application and specifying the data” on page 9-2 to open the Single-variable statistics input form and to load the data to analyze.
2. In the Col field, enter the number of the matrix column that holds the data you want to analyze.
3. In the Type field, press CHOOS and select the type of statistical data to measure:
   – Select SAMPLE if your data represents a sample of the population.
   – Select POPULATION if your data represents the entire population.
4. Place the cursor in a field for a statistic that you want to calculate and press CHK. Repeat for other statistics you want to calculate.
5. Press OK. The values you selected are calculated and displayed in a list in history.

The following single-variable statistics can be calculated:

Mean	Returns the arithmetic mean.
Std Dev	Returns the standard deviation.
Variance	Depending on the type that you selected, returns either the sample variance or the population variance.
Total	Returns the sum of the data.
Maximum	Returns the largest value in the data.
Minimum	Returns the smallest value in the data.

Generating frequencies

Frequency distributions describe how data is distributed across a specified set of sub-intervals, or bins. You specify:

- the minimum value for data elements to be included in the frequency distribution
- the bin number
- the bin size.

Starting from the minimum value, the statistics application sets up the number of intervals. Each interval is set to the size that you specify. From this, the statistics application defines the maximum value of data to be sampled.

To set up a frequency distribution for your data, use the following procedure:

1. Use the method described in “Starting an application and specifying the data” on page 9-2 to open the Frequencies input form and to load the data to analyze.
2. In the X-Min field, enter the minimum value for samples to be included in the analysis.
3. In the Bin Count field, enter the number of intervals, or bins.
4. In the Bin Width field, enter the size of each interval, or bin.

   The statistics application calculates the highest value to be included in the sample.
5. Press OK. The following data is returned in a list to history:
   - An array of integers, representing the number of data elements that fell into each interval, from lowest interval to highest interval.
   - A two-element vector—the first element represents the number of elements below the lowest value and the second element represents the number of elements above the highest allowable value.

Using statistics

Fitting a model to a set of data

You can use the statistics application to calculate Pearson's correlation coefficient for bivariate data. The statistics application quantifies the correlation between data in any two columns in a matrix. You can choose a regression model to apply to the data to find the relationship, or you can select the Best Fit option to allow the calculator to find the best correlation from its library of fit types.

The following four regression models are available for selection:

- Linear fit

  $y = b + mx$

- Logarithmic fit

  $y = b + m\ln x$

- Exponential fit

  $y = be^{mx}$ or $\ln y = \ln b + mx$

- Power fit

  $y = bx^{m}$ or $\ln y = \ln b + m\ln x$

To determine details of the regression model that applies to your data, use the following procedure:

1. Use the method described in "Starting an application and specifying the data" on page 9-2 to open the Fit Data input form and to load the data to analyze.
2. In the X-Col field, enter the column number that holds the independent variable values.
3. In the Y-Col field, enter the column number that holds the dependent variable values.
4. Place the cursor in the Model field and press CHOOS. A choose list containing the regression model options is displayed.
5. Select the regression model to apply to the data, or select Best fit to apply the model that best fits the data.
6. Press OK to calculate the regression details. The following items appear in history.
   - Item 1: the covariance value.
   - Item 2: the correlation coefficient.
   - Item 3: the regression formula.

Using statistics

Predicting a value based on the regression

Once you have performed a regression, you can use it to predict *y* values.

1. Follow steps 1 to 5 in the previous section to apply a regression to your data.
2. Press PRED. The Predict Values input form is displayed.
3. In the X field, enter the value for which you want to find the corresponding *y* value.
4. Move the cursor to the Y field and press PRED. The computed value, based on the regression, appears.

Although you can use this method to predict a value for *x* that corresponds to a known *y* value, the solution may not be mathematically correct.

Calculating summary statistics

You can use the summary statistics application to calculate up to six summary statistics on bivariate data.

To calculate summary statistics:

1. Use the method described in “Starting an application and specifying the data” on page 9-2 to open the Summary Statistics input form and to load the data to analyze.
2. In the X-Col and Y-Col fields, specify the columns that hold the data to analyze.
3. Use the arrow keys to navigate around the Calculate fields. Press CHK to choose the values that you want to calculate. A check mark appears against the ones you choose. The summary statistics that can be calculated are as follows:

ΣX	The sum of the values in the X-Col of ΣDAT.
ΣY	The sum of the values in the Y-Col of ΣDAT.
ΣX2	The sum of the squares of the values in the X-Col of ΣDAT.
ΣY2	The sum of the squares of the values in the Y-Col of ΣDAT.
ΣXY	The sum of the products of the X-Col and Y-Col pairs of ΣDAT
NΣ	The number of rows in ΣDAT.

4. Press OK to calculate the statistics. The statistics appear in history.

Plotting statistics

The following statistical plot types are available:

- Histogram
- Bar
- Scatter

By default, these plot types plot the data stored in ΣDAT. See chapter 4, “Plotting graphs”, for details on how to plot statistical data.

Inferential statistics

The inferential statistics capabilities of the HP 49G include calculation of confidence intervals and hypothesis tests based on the Normal Z-distribution or Student’s t-distribution.

Based on the statistics from one or two samples, you can test hypotheses and find confidence intervals for the following quantities:

- mean
- proportion
- difference between two means
- difference between two proportions

The calculator contains online help for each test. You access the online help by pressing HELP on the test input form.

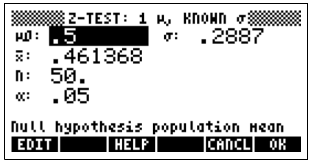
Example data

When you first access an input form for an inferential statistics test, by default the input form contains example data. This example data is designed to return meaningful results that relate to the test. It is useful for gaining an understanding of what the test does, and for demonstrating the test. The calculator’s online help provides a description of what the example data represents.

Using inferential statistics

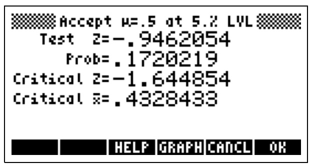
To use the inferential statistics functions:

1. Press ↱ STAT to access the Statistics menu.
2. Select the type of inferential statistics that you want.
   – Select HYPOTH. TESTS to display the hypothesis tests.
   – Select CONF. INTERVAL to display the confidence interval options.
3. From the list, select the hypothesis test or the confidence interval that you want. The input form, containing example data, is displayed. For example, when you select the Z-Test: 1 μ hypothesis test, the input form at the right appears.

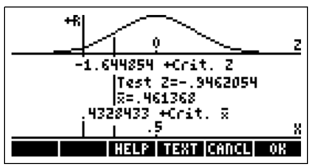


   For information on the test or confidence interval, and the example data, press HELP.
4. Enter your data, or leave the input form as it is to use the example data.
5. Press OK.
   – For hypothesis tests, a choose list appears, with the hypotheses to test against the null hypothesis. Select the hypothesis you want and press ENTER. The results of the test are displayed.
   – For confidence intervals, the results are displayed immediately.

   By default, the results are displayed in text format. The results for the one-sample Z-test example data are shown on the right.



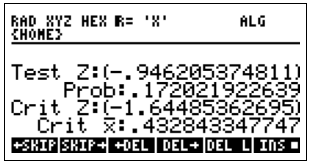
   To change input values, or to select a different hypothesis to test, press CANCEL to return to the previous screen.
6. Press GRAPH to display the results graphically. The results for the one-sample Z-test example data are shown on the right.



   To view the results in text format, press TEXT.

7. Press OK to close the inferential statistics application and return to the default screen. The results are copied to history.

   The results for the one-sample Z-test example data appear in history as shown on the right.



Hypothesis tests

You use hypothesis tests to test the validity of hypotheses that relate to the statistical parameters of one or two populations. The tests are based on statistics from samples of the population.

The HP 49G hypothesis tests use the Normal Z-distribution or Student's t-distribution to calculate probabilities.

One-Sample Z-Test

Menu name: Z–Test: 1 μ

On the basis of statistics from a single sample, measures the strength of the evidence for a selected hypothesis against the null hypothesis. The null hypothesis is that the population mean equals a specified value, μ_0

You select one of the following alternative hypotheses against which to test the null hypothesis:

$$H_1: \mu < \mu_0$$

$$H_2: \mu > \mu_0$$

$$H_3: \mu \neq \mu_0$$

Inputs

μ_0	Population mean given the null hypothesis.
σ	Population standard deviation.
$\bar{x}$	Sample mean.
n	Sample size.
α	Significance level.

Results

Test Z	Z-test statistic.
Prob	Probability associated with the Z-test statistic.
Critical Z	Boundary value of Z associated with the α level that you supplied.
Critical $\bar{x}$	Boundary value of $\bar{x}$ required by the α value that you supplied.

Two-Sample Z-Test

Menu name: Z–Test: μ1–μ2

On the basis of two samples, each from a separate population, measures the strength of the evidence for a selected hypothesis against the null hypothesis. The null hypothesis is that the mean of population 1 equals the mean of population 2: H_0: $\mu_1 = \mu_2$

You select one of the following alternative hypotheses against which to test the null hypothesis:

$$H_1{:}\mu_1 < \mu_2$$

$$H_2{:}\mu_1 > \mu_2$$

$$H_3{:}\mu_1 \neq \mu_2$$

Inputs

$\bar{x}$ 1	Sample 1 mean.
$\bar{x}$ 2	Sample 2 mean.
σ1	Population 1 standard deviation.
σ2	Population 2 standard deviation.
n1	Sample 1 size.
n2	Sample 2 size.
α	Significance level.

Results

Test Z	Z-test statistic.
Prob	Probability associated with the Z-test statistic.
Critical Z	Boundary value of Z associated with the α level that you supplied.

One-Proportion Z-Test

Menu name: Z-Test: 1 P

On the basis of statistics from a single sample, measures the strength of the evidence for a selected hypothesis against the null hypothesis. The null hypothesis is that the proportion of successes in the population equals a specified value, π_0

You select one of the following alternative hypotheses against which to test the null hypothesis:

$$H_1: \pi < \pi_0$$

$$H_2: \pi > \pi_0$$

$$H_3: \pi \neq \pi_0$$

Inputs

π_0	Population proportion of successes.
x	Number of successes in the sample.
n	Sample size.
α	Significance level.

Results

Test P	Proportion of successes in the sample.
Test Z	Z-test statistic.
Prob	Probability associated with the Z-test statistic.
Critical Z	Boundary value of Z associated with the α level that you supplied.

Two-Proportion Z-Test

Menu name: Z-Test: P1–P2

On the basis of statistics from two samples, each from a different population, measures the strength of the evidence for a selected hypothesis against the null hypothesis. The null hypothesis is that the proportion of successes in population 1 equals the proportion of successes in population 2: H_0: $\pi_1 = \pi_2$

You select one of the following alternative hypotheses against which to test the null hypothesis:

$$H_1 : \pi_1 < \pi_2$$

$$H_2 : \pi_1 > \pi_2$$

$$H_3 : \pi_1 \neq \pi_2$$

Using statistics

Inputs

X1	Sample 1 mean.
X2	Sample 2 mean.
n1	Sample 1 size.
n2	Sample 2 size.
α	Significance level.

Results

Test P1–P2	Difference between the proportions of successes in the two samples.
Test Z	Z-test statistic.
Prob	Probability associated with the Z-test statistic.
Critical Z	Boundary value of Z associated with the α level that you supplied.

One-Sample T-Test

Menu name: T-Test: 1 μ

The One-sample T-test is used when the population standard deviation is not known. On the basis of statistics from a single sample, measures the strength of the evidence for a selected hypothesis against the null hypothesis. The null hypothesis is that the sample mean has some assumed value: $H_0 : \mu = \mu_0$

You select one of the following alternative hypotheses against which to test the null hypothesis:

$$H_1 : \mu < \mu_0$$

$$H_2 : \mu > \mu_0$$

$$H_3 : \mu \neq \mu_0$$

Inputs

μ0	Population mean.
n	Sample size.
$\bar{x}$	Sample mean.
Sx	Sample standard deviation.
α	Significance level.

Results

Test T	T-test statistic.
Prob	Probability associated with the T-test statistic.
Critical T	Boundary value of T associated with the α level that you supplied.
Critical $\bar{x}$	Boundary value of $\bar{x}$ required by the α value that you supplied.

Using statistics

Two-Sample T-Test

Menu name: T-Test: µ1 – µ2

The Two-sample T-test is used when the population standard deviation is not known. On the basis of statistics from two samples, each sample from a different population, measures the strength of the evidence for a selected hypothesis against the null hypothesis. The null hypothesis is that the mean of population 1 equals the mean of population 2: H_0: $\mu_1 = \mu_2$

You select one of the following alternative hypotheses against which to test the null hypothesis

$$H_1{:}\mu_1 < \mu_2$$

$$H_2{:}\mu_1 > \mu_2$$

$$H_3{:}\mu_1 \neq \mu_2$$

Inputs

$\bar{x}$ 1	Sample 1 mean.
$\bar{x}$ 2	Sample 2 mean.
S1	Sample 1 standard deviation.
S2	Sample 2 standard deviation.
n1	Sample 1 size.
n2	Sample 2 size.
α	Significance level.
_Pooled?	Check this option to pool samples based on their standard deviations.

Results

Test T	T-test statistic.
Prob	Probability associated with the T-test statistic.
Critical T	Boundary value of T associated with the α level that you supplied.

Confidence intervals

The confidence interval calculations that the HP 49G can perform are based on the Normal Z-distribution or Student's t-distribution.

One-Sample Z-Interval

Menu name: Z–INT: 1 μ

This option uses the Normal Z-distribution to calculate a confidence interval for μ, the true mean of a population, when the true population standard deviation, σ, is known.

Inputs

$\bar{x}$	Sample mean.
σ	Population standard deviation.
n	Sample size.
C	Confidence level.

Results

Critical Z	Critical value for Z.
μ min	Lower bound for μ.
μ max	Upper bound for μ.

Two-Sample Z-Interval

Menu name: Z–INT: μ1– μ2

This option uses the Normal Z-distribution to calculate a confidence interval for the difference in the means of two populations, μ_1 and μ_2, when the population standard deviations, σ_1 and σ_2 are known.

Inputs

$\bar{x}$ 1	Sample 1 mean.
$\bar{x}$ 2	Sample 2 mean.
σ1	Population 1 standard deviation.
σ2	Population 2 standard deviation.
n1	Sample 1 size.
n2	Sample 2 size.
C	Confidence level.

Results

Critical Z	Critical value for Z.
$\Delta\mu$ Min	Lower bound for $\mu_1 - \mu_2$
$\Delta\mu$ Max	Upper bound for $\mu_1 - \mu_2$

One-Proportion Z-Interval

Menu name: Z–INT: 1 P

This option uses the Normal Z-distribution to calculate a confidence interval for the proportion of successes in a population for the case in which a sample of size, n, has a number of successes, x.

Inputs

x	Sample success count.
n	Sample size.
C	Confidence level.

Results

Critical Z	Critical value for Z.
π Min	Lower bound for π.
π Max	Upper bound for π.

Two-Proportion Z-Interval

Menu name: Z–INT: P1 – P2

This option uses the Normal Z-distribution to calculate a confidence interval for the difference in the proportions of successes in two populations.

Inputs

x1	Sample 1 success count.
x2	Sample 2 success count.
n1	Sample 1 size.
n2	Sample 2 size.
C	Confidence level.

Results

Critical Z	Critical value for Z.
$\Delta\pi$ Min	Lower bound for the difference in proportions of successes.
$\Delta\pi$ Max	Upper bound for the difference in proportions of successes.

One-Sample T-Interval

Menu name: T–INT: 1 μ

This option uses the Student's t-distribution to calculate a confidence interval for μ, the true mean of a population, for the case in which the true population standard deviation, σ, is unknown.

Inputs

$\bar{x}$	Sample mean.
Sx	Sample standard deviation.
n	Sample size.
C	Confidence level.

Results

Critical T	Critical value for T.
μ Min	Lower bound for μ.
μ Max	Upper bound for μ.

Two-Sample T-Interval

Menu name: T–INT: $\mu 1 - \mu 2$

This option uses the Student's t-distribution to calculate a confidence interval for the difference in the means of two populations, $\mu_1 - \mu_2$, when the population standard deviations, $\sigma_1$and σ_2, are unknown.

Inputs

$\bar{x}$ 1	Sample 1 mean.
$\bar{x}$ 2	Sample 2 mean.
s1	Sample 1 standard deviation.
s2	Sample 2 standard deviation.
n1	Sample 1 size.
n2	Sample 2 size.
C	Confidence level.
_Pooled	Whether or not to pool the samples based on their standard deviations.

Results

Critical T	Critical value for T.
$\Delta\mu$ Min	Lower bound for $\mu_1 - \mu_2$.
$\Delta\mu$ Max	Upper bound for $\mu_1 - \mu_2$.