

Revision section: the basics

R1.3 Confidence intervals: a way of precisely representing uncertainty

A whole range of descriptive statistics can be produced by SAS for a variable. Below we give an example for the exam results described in Box R1.1.

| SAS COMMANDS FOR BOX R1.1 Descriptive statistics and confidence intervals for single samples | |
|---|--|
| Commands | <pre>proc capability data=gandh.revision cibasic trimmed=0.1; var EXAMRES; histogram / normal; run;</pre> |
| Menu route | Statistics > Descriptive > Distributions EXAMRES → Analysis <input type="checkbox"/> Plots <input checked="" type="checkbox"/> Histogram Statistics > Hypothesis Tests > One-Sample t-test for a Mean EXAMRES → Variable <input type="checkbox"/> Tests with Confidence Intervals tab <input checked="" type="radio"/> Interval |

SAS OUTPUT FOR BOX R1.1 **Descriptive statistics and confidence intervals for single samples**

The CAPABILITY Procedure
Variable: EXAMRES

Moments

| | | | |
|-----------------|------------|------------------|------------|
| N | 30 | Sum Weights | 30 |
| Mean | 50.333333 | Sum Observations | 1510 |
| Std Deviation | 15.8556709 | Variance | 251.402299 |
| Skewness | -0.1818838 | Kurtosis | -0.5774078 |
| Uncorrected SS | 83294 | Corrected SS | 7290.66667 |
| Coeff Variation | 31.5013329 | Std Error Mean | 2.8948362 |

Basic Statistical Measures

| Location | | Variability | |
|----------|----------|---------------------|-----------|
| Mean | 50.33333 | Std Deviation | 15.85567 |
| Median | 50.50000 | Variance | 251.40230 |
| Mode | 28.00000 | Range | 65.00000 |
| | | Interquartile Range | 26.00000 |

NOTE: The mode displayed is the smallest of 5 modes with a count of 2.

Basic Confidence Limits Assuming Normality

| Parameter | Estimate | 95% Confidence Limits | |
|---------------|----------|-----------------------|----------|
| Mean | 50.33333 | 44.41273 | 56.25394 |
| Std Deviation | 15.85567 | 12.62757 | 21.31502 |
| Variance | 251.4023 | 159.4554 | 454.33 |

Tests for Location: Mu0=0

| Test | -Statistic- | -----p-value----- | |
|-------------|-------------|-------------------|--------|
| Student's t | t 17.38728 | Pr > t | <.0001 |
| Sign | M 15 | Pr >= M | <.0001 |
| Signed Rank | S 232.5 | Pr >= S | <.0001 |

Trimmed Means

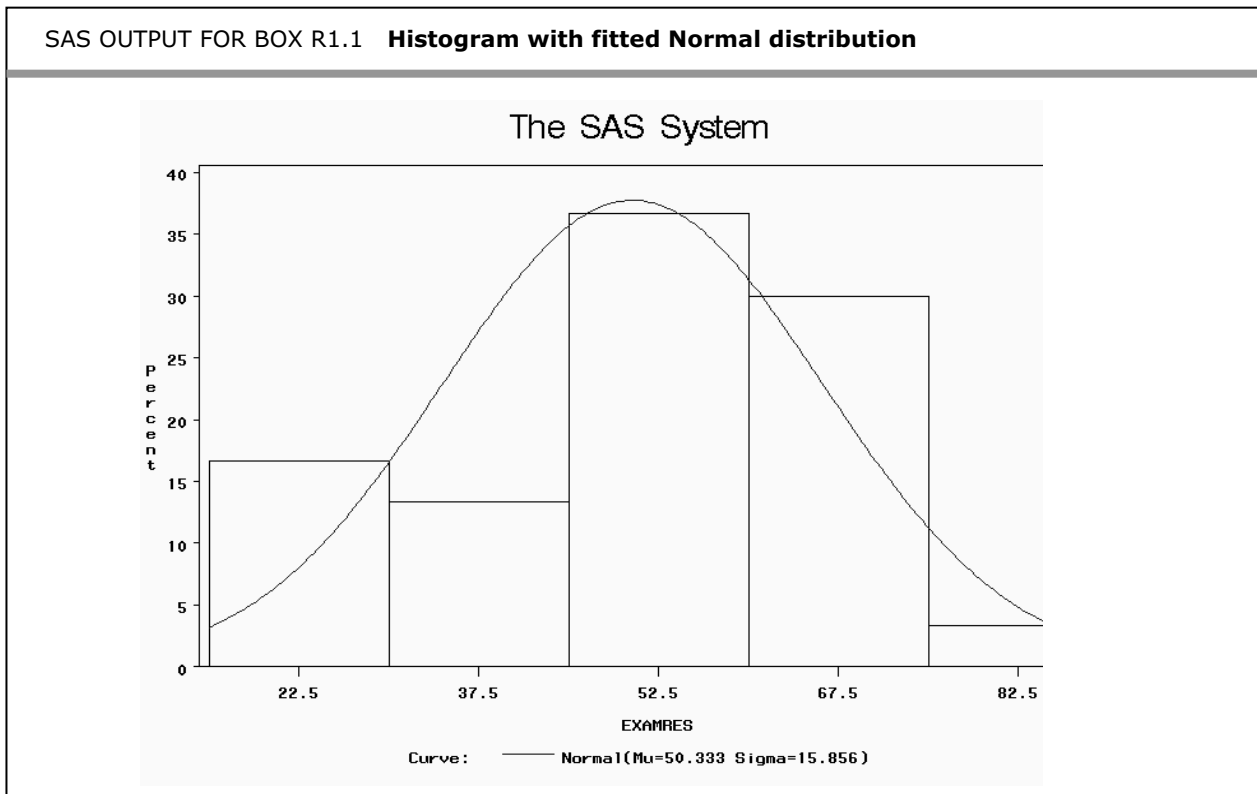
| Percent Trimmed in Tail | Number Trimmed in Tail | Trimmed Mean | Std Error Trimmed Mean | 95% Confidence Limits | DF | t for H0: Mu0=0.00 | Pr > t |
|-------------------------|------------------------|--------------|------------------------|-----------------------|----|--------------------|---------|
| 10.00 | 3 | 50.70833 | 3.192458 | 44.10423 57.31244 | 23 | 15.88379 | <.0001 |

Quantiles (Definition 5)

| Quantile | Estimate |
|------------|----------|
| 100% Max | 83.0 |
| 99% | 83.0 |
| 95% | 71.0 |
| 90% | 67.5 |
| 75% Q3 | 65.0 |
| 50% Median | 50.5 |
| 25% Q1 | 39.0 |
| 10% | 28.0 |
| 5% | 25.0 |
| 1% | 18.0 |
| 0% Min | 18.0 |

Extreme Observations

| ----Lowest---- | | ----Highest--- | |
|----------------|-----|----------------|-----|
| Value | Obs | Value | Obs |
| 18 | 26 | 67 | 8 |
| 25 | 13 | 67 | 24 |
| 28 | 29 | 68 | 30 |
| 28 | 15 | 71 | 4 |
| 29 | 18 | 83 | 2 |



SAS produces a wealth of information in addition to the confidence interval. However, you can see that the 95 percent confidence interval is the same interval as that calculated in the main text. SAS has also drawn a Normal distribution with a mean and variance corresponding to our sample. Confidence intervals and Normal distributions are covered in the revision section of the main text.

R1.4 The null hypothesis—taking the conservative approach

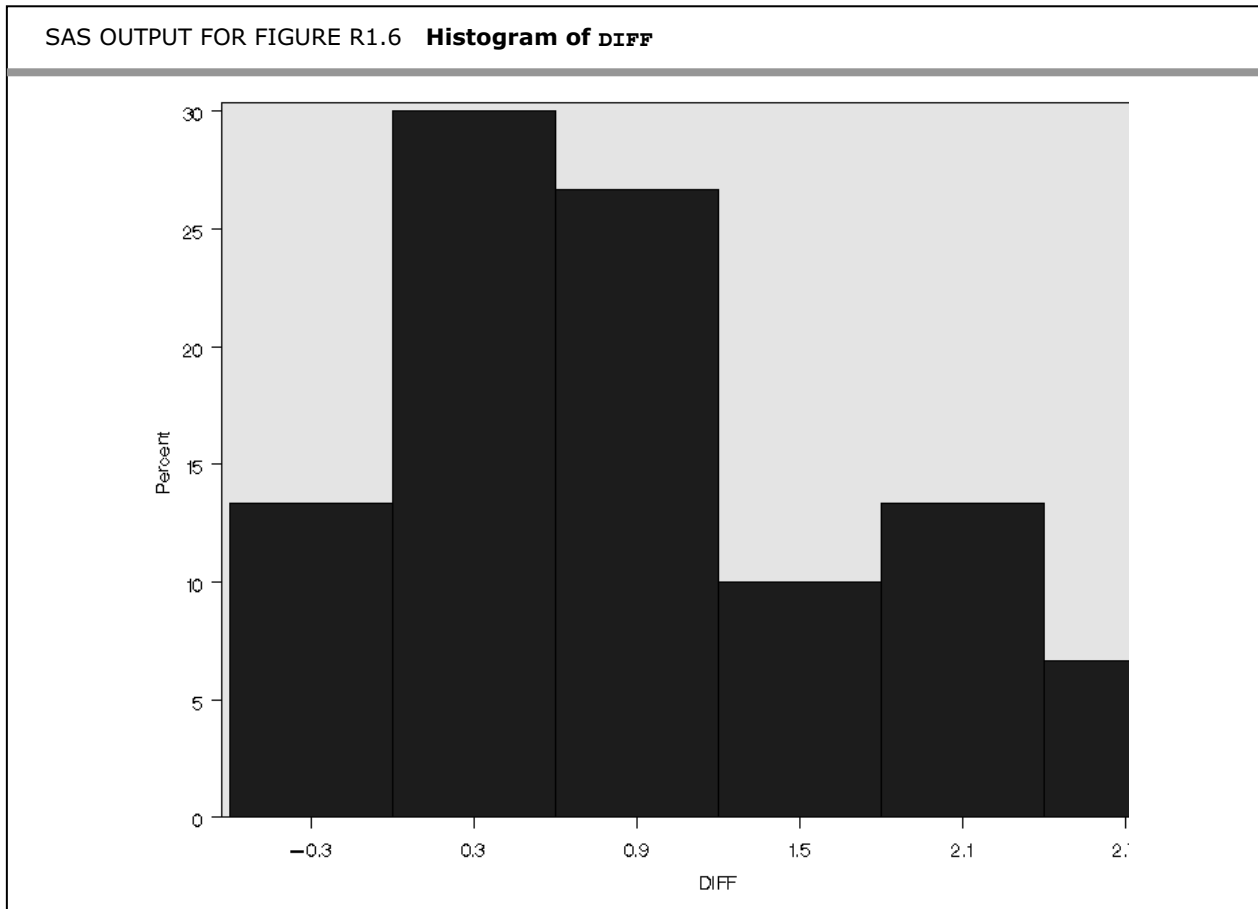
In this section, a one sample t -test is conducted on the data variable DIFF. First, a histogram of the data is plotted.

SAS COMMANDS FOR FIGURE R1.6

```
Commands  proc capability noprint;
           var DIFF;
           histogram;
           run;
```

Menu route Graphs > Histogram

DIFF → Analysis



Then the one sample t -test is conducted. The test result by the command line route will be found among the copious output, in the section headed 'Tests for Location: $\mu_0=0$ '. The menu route output is briefer and is shown here.

| SAS COMMANDS FOR BOX R1.2 One sample t-test | |
|---|--|
| Commands | <pre>proc capability; var DIFF; run;</pre> |
| Menu route | Statistics > Hypothesis Tests > One-Sample t -test for a Mean DIFF → Variable 0 → Mean |

SAS OUTPUT FOR BOX R1.2 **One sample *t*-test**

```

One Sample t-test for a Mean

Sample Statistics for DIFF

N          Mean      Std. Dev.    Std. Error
-----
30         0.86       0.84         0.15

Hypothesis Test

Null hypothesis: Mean of DIFF = 0
Alternative:     Mean of DIFF ^= 0

t Statistic   Df      Prob > t
-----
5.636         29     <.0001

```

R1.5 Comparing two means**Two sample *t*-test**

The two sample *t*-test is conducted on the *squirrels* dataset. First the data are transformed to natural logarithms.

SAS COMMANDS TO TRANSFORM DATA

```

Commands  data;
          set gandh.revision;
          LMALE =log(MALE);
          LFEM =log(FEMALE);
          run;

```

```

Menu route  Edit > Mode > Edit...

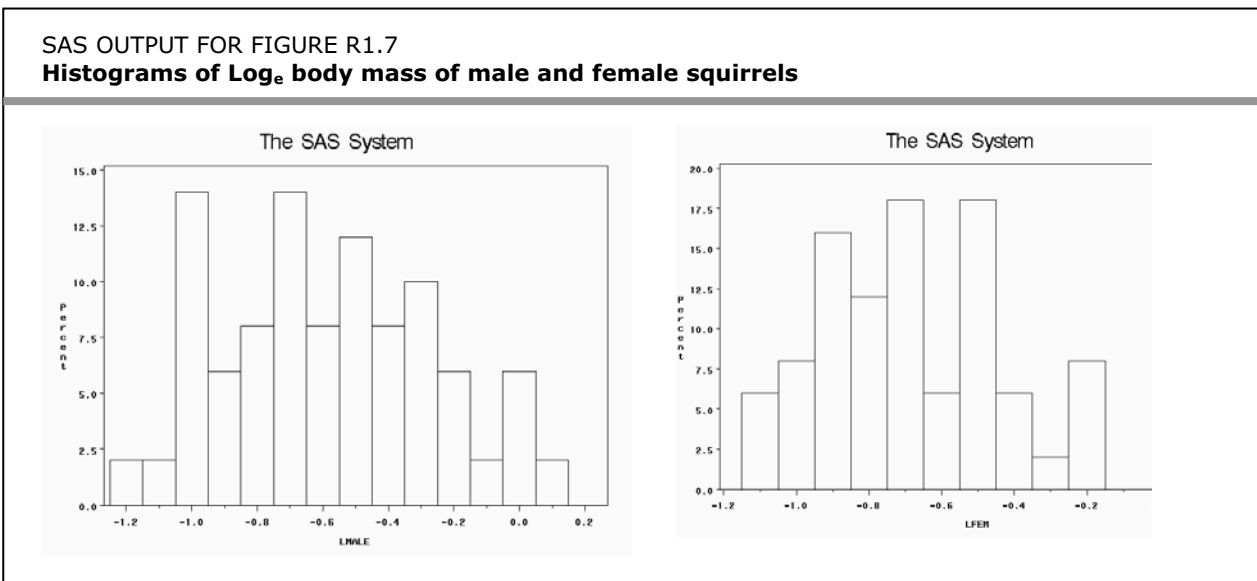
           Data > Transform > Compute...
           Type "LMALE" into top left box
           log(MALE) → Main pane

           Data > Transform > Compute...
           Type "LFEM" into top left box
           log(FEMALE) → Main pane

```

If you use the command line route, you will find a new dataset in the work library called data1. This contains the original dataset, and the transformed variables in addition. Both the raw data and the logged data are plotted—but here we just show how to do this for the logged data. The midpoint/cutpoint command allows you to specify the bar divisions, and so ensure the two histograms are plotted on the same X scale.

| SAS COMMANDS FOR FIGURE R1.7 Histograms of Log_e body mass of male and female squirrels | |
|---|--|
| Commands | <pre>proc capability noprint; var LMALE LFEM; histogram / midpoints = -1.2 to 0.1 by 0.1; run;</pre> |
| Menu route | <p>Graphs > Histogram...</p> <p>LMALE LFEM → Analysis</p> <p>Type -1.3, 0.1, 0.1 into: Midpoints of histogram intervals</p> |



Finally, the two sample *t*-test. SAS uses the ‘pooled’ method to calculate the standard errors in the command route. In the menu route, whose output is shown, the *t*-test is given twice. Once ‘if variances are equal’ (which is what we have called ‘pooled’) and again ‘if variances are not equal’ (which is the unpooled method, in which you will find non-integer degrees of freedom). The pooled method assumes that one error variance can be estimated jointly for the two samples.

SAS COMMANDS FOR BOX R1.6 **Two sample t-test**

```

Commands  data;
           set gandh.revision;
           LMALE=log(MALE);
           LFEM=log(FEMALE);
proc ttest;
           var LMALE LFEM;
run;

```

Menu route Statistics > Hypothesis Tests > Two-Sample t-test for Means

Two variables

LMALE → Group 1

LFEM → Group 2

SAS OUTPUT FOR BOX R1.6 **Two sample t-test**

Two Sample t-test for the Means of LMALE and LFEM

Sample Statistics

| Group | N | Mean | Std. Dev. | Std. Error |
|-------|----|----------|-----------|------------|
| LMALE | 50 | -0.57858 | 0.3248 | 0.0459 |
| LFEM | 50 | -0.68781 | 0.2453 | 0.0347 |

Hypothesis Test

Null hypothesis: Mean 1 - Mean 2 = 0
 Alternative: Mean 1 - Mean 2 \neq 0

| If Variances Are | t statistic | Df | Pr > t |
|------------------|-------------|-------|--------|
| Equal | 1.898 | 98 | 0.0607 |
| Not Equal | 1.898 | 91.17 | 0.0609 |

One and two tailed tests

SAS's PROC TTEST will not perform a one-tailed test, and so we give only the menu route here. To convert a one-tailed to a two-tailed test requires an additional checkbox, as shown below:

SAS COMMANDS FOR BOX R1.9
A one-tailed two sample t-test of the logs of squirrel body masses

Menu route Statistics > Hypothesis Tests > Two-Sample t-test for Means...

Two variables

LMALE → Group 1

LFEM → Group 2

Alternative: Mean 1 - Mean 2 > 0

SAS OUTPUT FOR BOX R1.9
A one-tailed two sample t-test of the logs of squirrel body masses

Two Sample t-test for the Means of LMALE and LFEM

Sample Statistics

| Group | N | Mean | Std. Dev. | Std. Error |
|-------|----|----------|-----------|------------|
| LMALE | 50 | -0.57858 | 0.3248 | 0.0459 |
| LFEM | 50 | -0.68781 | 0.2453 | 0.0347 |

Hypothesis Test

Null hypothesis: Mean 1 - Mean 2 <= 0
Alternative: Mean 1 - Mean 2 > 0

| If Variances Are | t statistic | Df | Pr > t |
|------------------|-------------|-------|--------|
| Equal | 1.898 | 98 | 0.0303 |
| Not Equal | 1.898 | 91.17 | 0.0305 |

In this case we chose the alternative hypothesis that 'Mean 1–Mean 2 > 0' because our alternative hypothesis was that the LMALE mean was greater than the LFEM mean. However, we could just as easily have had the alternative hypothesis that the LMALE mean was less than the LFEM mean. However, having noted the direction of the difference, (LMALE > LFEM), then the test would most definitely not be significant in the other direction! Our one-tailed test has converted a non-significant *p*-value into a significant one—but see main text for a health warning!