

**Lab 12: t-tests and confidence intervals**  
**STT 421: Summer, 2004**  
**Vince Melfi**

There are *many* ways to perform t-tests and compute confidence intervals for means in SAS. We'll investigate a few of these in today's lab, starting by using SAS as a glorified t-table and then looking at more automated methods.

## 12.1 Computing by hand

Exercise 7.29 contains data on the accuracy of radon detectors. Specifically, 12 home-use radon detectors were placed in a chamber that exposed them to 105 picocuries per liter of radon. The 12 detectors readings were

```
91.9  97.8 111.4 122.3 105.4 95.0
103.8 99.6 96.6  119.3 104.8 101.7
```

First we'll use SAS to compute a 97% confidence interval for the mean reading  $\mu$  of all detectors of this type. The formula is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}},$$

where  $t^*$  is the 98.5th percentile of the  $t$  distribution with  $n - 1 = 11$  degrees of freedom. (We need 97% of the area under the  $t$  density to be between  $-t^*$  and  $t^*$ , which means that 1.5% of the area is to the right of  $t^*$ , which means that 98.5% of the area is to the left of  $t^*$ .) The only new command is **tin**v, which gives the percentiles of the  $t$  distribution. First we read in and print the data.

```
data radon;
  input reading;
  cards;
  91.9
  97.8
  111.4
  122.3
  105.4
  95.0
  103.8
  99.6
  96.6
  119.3
  104.8
  101.7
  ;

proc print data = radon;
run;
```

Now we use **proc univariate** to compute the mean and standard deviation, and use the function **tinvt** to compute  $t^*$ . Note also that the **keep** statement tells SAS which variables to keep in a dataset.

```
proc means data = radon;
  output out = radonstats mean = xbar stddev = s;

data radonci;
  set radonstats;
  tstar = tinvt(.985, 11);
  lower = xbar - tstar * (s/sqrt(12));
  upper = xbar + tstar * (s/sqrt(12));
  keep xbar s tstar lower upper;

proc print data = radonci;

run;
```

1. What is the 97% confidence interval?

Next we'll use SAS to help us perform a test of  $H_0: \mu = 115$  versus  $H_a: \mu \neq 115$ . We'll use some of what we created above. Make sure you can figure out how I computed the p-value. Drawing a picture might help.

```
data radontest;
  set radonci;
  tstat = (xbar - 105)/(s/sqrt(12));
  pval = 2 * cdf('t', -abs(tstat), 11);
  keep tstat pval;

proc print data = radontest;

run;
```

1. What is the value of the  $t$  statistic?
2. What is the p-value of the test?

## 12.2 Using proc means

It's possible to compute confidence intervals and perform t-tests more automatically using **proc means**.<sup>1</sup> For confidence intervals, this requires asking for the lower confidence limit `lclm` and the upper confidence limit `uclm` and specifying the confidence level via the keyword `alpha`. Then SAS computes a  $1 - \alpha$  level interval. For example, the following SAS code requests a 97% interval. You should get the same answer as above.

---

<sup>1</sup>It's also possible to do this using **proc univariate** and **proc ttest**. We'll see how to use **proc ttest** below.

```
proc means data = radon lclm uclm alpha = .03;

run;
```

We have to be a bit more clever to get SAS to perform the t-test automatically, because SAS only likes to test the null hypothesis  $H_0: \mu = 0$ . To trick SAS into testing our null hypothesis, we will first create a new dataset which contains our data minus 105. Testing  $H_0: \mu = 0$  for this new data is equivalent to testing  $H_0: \mu = 105$  for the original data. Requesting `t` and `probt` tells SAS that we'd like the t-statistic and the p-value.

```
data radonminus105;
  set radon;
  readingminus105 = reading - 105;
  drop reading;

proc means data = radonminus105 t probt;

run;
```

1. What is the value of the  $t$  statistic?
2. What is the p-value?
3. By default SAS computes the p-value for a two-sided alternative hypothesis  $H_a: \mu \neq 105$ . Use the output from above to compute the p-value for the test of  $H_a: \mu < 105$ .

## 12.3 Paired data

Remember that to test hypotheses about the mean difference for paired data, we apply the procedure for testing hypotheses about the mean of a population to the difference of the observations. So we can use the same techniques as above for paired data.

1. Use `proc means` to perform the test specified in Exercise 7.35.

## 12.4 Comparing two samples; `proc ttest`

As you know from the text, there are various methods for performing a t test to compare the means of two populations based on independent samples from the populations. Many of these methods are available in SAS through `proc ttest`, which also performs one-sample t tests.

We'll start by seeing how to perform the test of  $H_0: \mu = 115$  versus  $H_a: \mu \neq 115$  and to compute a confidence interval for the radon data using `proc ttest`.

```
proc ttest data=radon h0=105 alpha = .03;

run;
```

We'll illustrate the use of **proc ttest** to compare means via independent samples using the DRP data from Table 7.2 in the text. These data give the Degree of Reading Power scores for third graders. There are 21 students who take part in new directed reading activities (the treatment group) and 23 students who do not (the control group). The data are available in the file `u:\msu\course\stt\421\su2001\textdata\Chap07\Ta07_002.dat`. The file contains the observation number, a character variable indicating whether the student was in the treatment or control group, a numerical variable indicating whether the student was in the treatment or control group, and the actual scores. Here is the SAS code to read in and print the data.

```
data drp;
  infile 'u:\msu\course\stt\421\su2001\textdata\Chap07\Ta07_002.dat';
  input obsnum group $ codedgroup score;

proc print data=drp;

run;
```

We can use either the character treatment indicator `group` or the numerical treatment indicator `codedgroup` in **proc ttest**. We'll opt for the former below. Here is the SAS code; again, `1 - alpha` is the confidence level for confidence intervals.

```
proc ttest data=drp alpha = .01;
  class group;
  var score;

run;
```

### 12.4.1 Explanation of output

1. Under the heading **Statistics** we get basic statistics for the two groups and the difference, including confidence intervals for the confidence level we've chosen.
2. Under the heading **T-tests** we get the test statistics, degrees of freedom, p-values, etc. for two types of tests: The pooled test, which is valid when the two populations to be compared have equal variances, and the Satterthwaite test, which does not require this assumption. Note that by default these test  $H_0: \mu_1 = \mu_2$  versus  $H_a: \mu_1 \neq \mu_2$ . As above, you should be able to use the output to compute p-values for one-sided alternative hypotheses.
3. We'll ignore the output under the heading **Equality of Variances**.
4. Question: What is the test statistic and p-value for the Pooled t-test? Would you reject  $H_0$  at the level  $\alpha = 0.01$ ? Would you reject  $H_0$  at the level  $\alpha = 0.05$ ?
5. Question: What is the test statistic and p-value for the Satterthwaite t-test? Would you reject  $H_0$  at the level  $\alpha = 0.01$ ? Would you reject  $H_0$  at the level  $\alpha = 0.05$ ?
6. Question: Is there much difference between the two procedures in this case?