STAT:2010/4200, Statistical Methods and Computing

Instructor: Cowles

Lab 4, Mar. 21, 2015

t-tests

Please find the following datasets on the course web page:

1. iowacorn.dat
2. poverty.dat

SAS for one-sample t-tests

- SAS automatically does a two-sided test

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

We will use the "iowacorn.dat" data on annual precipitation, temperature, corn production, and acres harvested in corn in each year of a 10 year period, to test two sets of hypotheses. The first is for annual precipitation in inches:

$$H_0 : \mu_{precip} = 35$$
$$H_a : \mu_{precip} \neq 35$$

The second is for annual average temperature:

$$H_0 : \mu_{temp} = 45$$
$$H_a : \mu_{temp} \neq 45$$

We will test both null hypotheses at the .10 significance level. We will first use the confidence-interval method.

```
data corn ;
* infile 'c:\temp\iowacorn.dat' ;
input precip temp corn acres ;
datalines ;
* note: copy and paste data in here ;
;
run ;

proc means data = corn n mean stddev stderr clm alpha = .10 ;
var precip temp ;
run ;
```


Variable	N	Mean	Std Dev	Std Error	Lower 90% CL for Mean
precip	10	33.9100000	6.6430331	2.1007115	30.0591585
temp	10	48.7610000	1.6182463	0.5117344	47.8229330

Variable	Upper 90% CL for Mean
precip	37.7608415
temp	49.6990670

1. What assumptions are necessary to justify computing t confidence intervals and using them to do t hypothesis tests?

2. The first confidence interval that SAS produced above was (30.06, 37.76). We are 90% confident that `______` is in this interval.

3. What can we conclude from this confidence interval about the hypothesis test regarding μ_{precip}?

4. What can we conclude from the other confidence interval about the hypothesis test regarding μ_{temp}?

One-sample t-tests using proc univariate

Proc univariate knows how to do only one kind of t-test:

- one-sample
- two-sided

```
proc univariate mu0 = 35 45 data = corn ;
var precip temp ;
run ;
```

```
            The UNIVARIATE Procedure
                Variable: precip


           Tests for Location: Mu0=35

Test           -Statistic-    -----p Value------

Student's t    t  -0.51887    Pr > |t|    0.6164
Sign           M         0    Pr >= |M|   1.0000
Signed Rank    S      -2.5    Pr >= |S|   0.8457
```

```
            The UNIVARIATE Procedure
                Variable:  temp


           Tests for Location: Mu0=45

Test           -Statistic-    -----p Value------

Student's t    t  7.349515    Pr > |t|    <.0001
Sign           M         5    Pr >= |M|   0.0020
Signed Rank    S      27.5    Pr >= |S|   0.0020
```

Paired t-test

To carry out the hypothesis test of interest, we apply one-sample procedures to the *differences* between values measured on members of each pair.

We are interested in whether life expectancy is the same for males as for females. We have a dataset containing various demographic and public health variables on 97 countries in the world in the early 1990s. Two variables reported on each country are the life expectancy at birth for males and the life expectancy at birth for females.

Our null hypothesis is that the mean life expectancy for males in the population of all countries in the world is the same as the mean life expectancy for females in the population of all countries.

We will do a two-sided test, because we do not know in advance whether to expect μ_1 (mean male life expectancy) to be higher or lower than μ_2 (mean female life expectancy).

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

or equivalently:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

or equivalently:

$$H_0 : \delta = 0$$
$$H_a : \delta \neq 0$$

where δ denotes $\mu_1 - \mu_2$.

We will use the *observed differences* between the male and female life expectancies observed on each country as our data to to carry out the hypothesis test regarding δ at the .05 significance level.

Note that by default, `proc univariate` tests the null hypothesis that $\mu = 0$, so in this case we don't have to give it a value for `mu0`.

```
data poverty ;
* infile 'c:\temp\poverty.dat' ;
length country $20. ;
input livebrth death infdeath mlifeexp flifeexp pcgdp group @53 country ;
diff = mlifeexp - flifeexp ;
datalines ;
* note: copy and paste data in here ;
;
run ;

proc means data = poverty n mean stddev stderr clm alpha = .05 ;
var diff ;
run ;

proc univariate data = poverty ;
var diff ;
run ;
```

The MEANS Procedure

Analysis Variable : diff

N	Mean	Std Dev	Std Error	Lower 95% CL for Mean	Upper 95% CL for Mean
97	-4.6655670	2.3711209	0.2407509	-5.1434537	-4.1876803

The UNIVARIATE Procedure
Variable: diff

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	-19.3792	Pr > |t|	<.0001
Sign	M	-44.5	Pr >= |M|	<.0001
Signed Rank	S	-2352	Pr >= |S|	<.0001