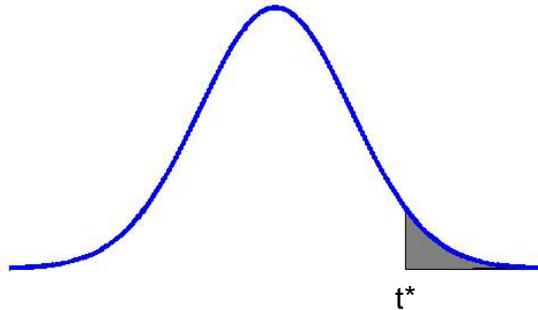


The t-model

STATA can be used to make calculations regarding the probabilities of the right tail of the t-model, using the commands `ttail` and `invttail`. This can be used to obtain critical values for confidence intervals and hypothesis tests, as well as p-values.



If you know t^* and want to calculate the area above it under the t-model with df degrees of freedom (shown in gray), use the command `ttail(df, t*)`.

If you know the area in gray, α (e.g. 5%), and want to calculate t^* , use the command `invttail(df,alpha)`.

Ex. 1 Suppose you want to calculate the critical value of t for a 90% confidence interval with 17 degrees of freedom, i.e. you want to find the value of t^* for which 5% of the area under the curve lies above t^* and 5% lies below $-t^*$. To find this value using STATA type:

```
. display invttail(17,0.05)
```

in the STATA command window. This gives us the 95th percentile of the t-model with 17 degrees of freedom, which corresponds to the critical value for a 90% confidence interval. In the Results window the value `1.7396067` is shown (Compare this value with the one given by the table in the back of the book).

Ex. 2 Suppose we want to find the p-value for $t \geq 2.09$ with 4 degrees of freedom. To find this value using STATA type:

```
. display ttail(4,2.09)
```

in the STATA command window. This gives us the value `0.05241536`, which corresponds to our p-value.

Note that the p-value for $t \leq 2.09$ (the area to the left of 2.09) with 4 degrees of freedom, would be given by

```
.display 1-ttail(4,2.09)
```

The p-value for $|t| \geq 2.09$ (two-sided test) with 4 degrees of freedom, would be given by

```
.display 2*ttail(4,2.09)
```

Confidence Intervals and T-Tests

The *Cars* data set gives the price in dollars and the weight in pounds for a number of 1991 model four-door sedans listed in a particular auto guide. American made cars are coded with a "0" and foreign brands are coded with a "1". The car data can be accessed by typing:

use <http://www.stat.columbia.edu/~martin/W1111/Data/Cars>

in the STATA command window.

To construct a level C confidence interval for the variable *var*, use the command

`ci var, level(C)`

For example, to get a 90% CI for the average price among all cars use the command

`ci price, level(90)`

This gives the following output:

Variable	Obs	Mean	Std. Err.	[90% Conf. Interval]	
price	84	16705.01	938.9414	15143.16	18266.87

Reading this output we see that a 90% CI for the average price is (15143.16, 18266.87).

To construct either a one or two-sample t-test use the command `ttest`. For example suppose the null hypothesis of our test is that the mean price of all four-door sedans is equal to \$18,000 and the alternative hypothesis is that the mean is less than \$18,000. To investigate this claim we need to use a [one-sample t-test](#). This can be done using the command

`ttest price = 18000`

This gives the following output:

One-sample t test					
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
price	84	16705.01	938.9414	8605.54	14837.5 18572.53
Degrees of freedom: 83					
Ho: mean(price) = 18000					
Ha: mean < 18000		Ha: mean != 18000		Ha: mean > 18000	
t = -1.3792		t = -1.3792		t = -1.3792	
P < t = 0.0858		P > t = 0.1715		P > t = 0.9142	

This command gives summary statistics for the variable `price` as well as the results of three different tests of significance that correspond to each of the possible alternative hypothesis.

For our test, we need to look under the column that reads `Ha: mean<18000`, as this corresponds to the alternative hypothesis we stated above. The P-value of this particular test is equal to 0.0858.

Suppose instead we want know whether there is a significant difference between the mean price of foreign and domestic four-door sedans. To investigate this claim we need to conduct a `two-sample t-test` to compare the mean price of the foreign and domestic cars.

To perform this test, use the command:

```
ttest price, by(cartype) unequal
```

which gives the following output:

Two-sample t test with unequal variances						
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	41	15744.66	1102.973	7062.471	13515.47	17973.85
1	43	17620.7	1502.597	9853.189	14588.33	20653.06
combined	84	16705.01	938.9414	8605.54	14837.5	18572.53
diff		-1876.039	1863.96		-5588.264	1836.186

Satterthwaite's degrees of freedom: 76.2196

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
t = -1.0065	t = -1.0065	t = -1.0065
P < t = 0.1587	P > t = 0.3174	P > t = 0.8413

Again, this command gives summary statistics as well as the results of three different tests of significance corresponding to each of the possible alternative hypothesis. For example, if our alternative hypothesis was that there is a difference between the means, the corresponding P-value is equal to 0.3174.

HOMEWORK:

Q1. Do problem 23.2 in the textbook.

Q2. Answer the following questions about the *Cars* data set described above.

1. Read the car data by typing:

use [http://www.stat.columbia.edu/~martin/W1111 /Data/Cars](http://www.stat.columbia.edu/~martin/W1111/Data/Cars)

in the STATA command window.

2. Construct a 95% CI for the average weight of all four-door sedans.

3. Is there significant evidence that the mean weight of all four-door sedans is below 3,100 pounds?

(a) State the appropriate null and alternative hypothesis.

(b) What is the P-value of the test?

(c) Are the results significant at the 5% level?

4. Is there a significant difference in weight between foreign and domestic cars?

(a) State the appropriate null and alternate hypothesis.

(b) What is the P-value of the test?

(c) Is there a significant difference between the weights at the 5% level?

Hand in your log file together with the answers to the questions above.