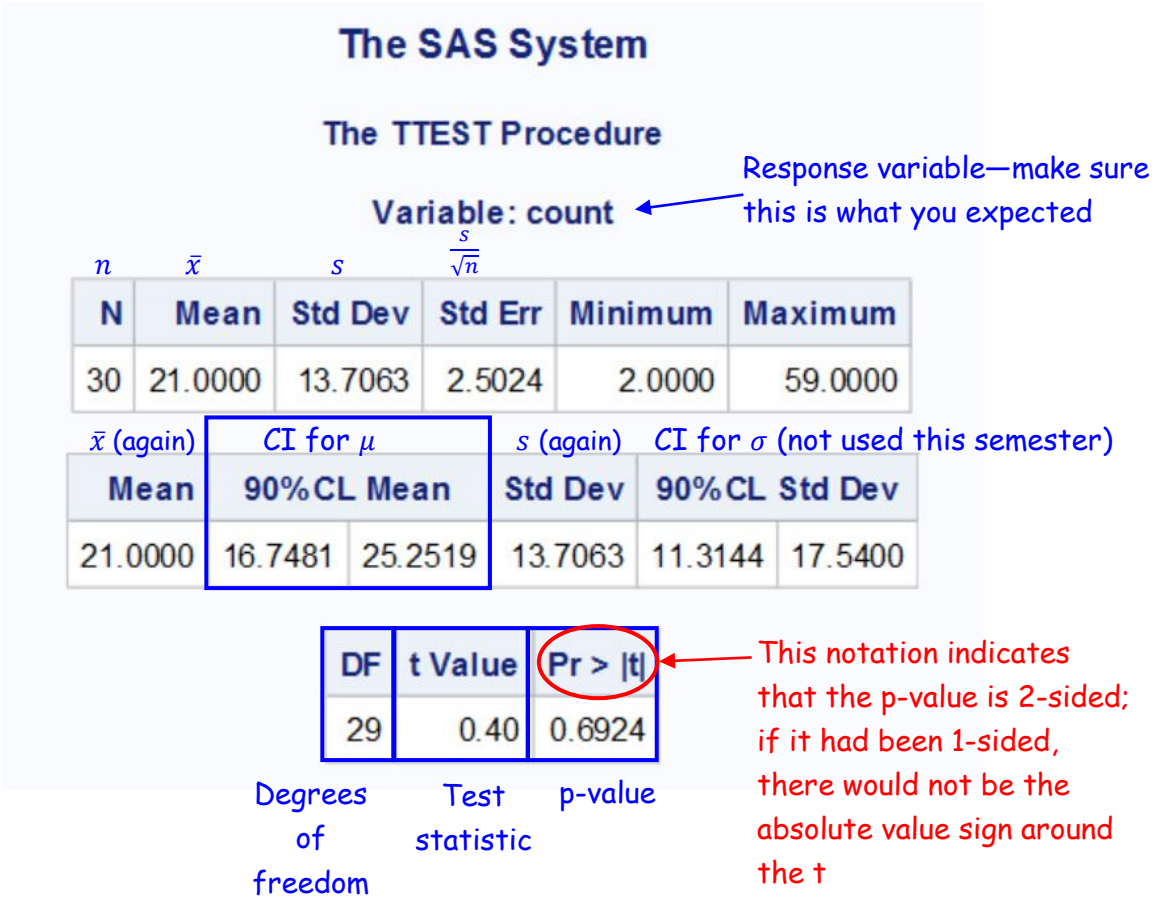
SAS CODE USING DATA FOR LAUGHTER EXAMPLE ON PAGE 10 OF NOTE OUTLINE 5:

```
* Note: text appearing in green is a comment;
* It can be used to explain code but is ignored by the computer when running code;
* Shorter comments (like these) can be placed between an asterisk (*) and a semi-colon;
* Longer comments can be placed between these marks: /* COMMENT HERE */


* All SAS programs start with a DATA step to read in the data;
* Additional data steps could be used to prepare data for analysis if necessary;
* Note that each line is ended with a semi-colon;
* An explanation of the code follows the DATA step;

DATA laughs;
input count;
datalines;
8
28
31
3
29
10
2
42
24
9
21
33
17
26
28
19
59
12
8
45
11
28
27
6
26
8
24
3
32
11
;
run;


/* Explanation of the DATA step:
> The INPUT statement tells SAS the names of the variable(s).
  If any categorical variables are present, their name should followed by a dollar sign,
  e.g. INPUT name $;
> The DATALINES statement tells SAS that data will be entered directly--this is the
  easiest way to read in small datasets. Larger datasets can be read in using the IMPORT
  DATA option under the FILE menu.
> The 30 values for this dataset follow the DATALINES command.
> The RUN statement tells SAS to execute the previous code.  */
```

```
* Analyzing data occurs in a procedure step, or PROC;
* PROC TTEST is used for both confidence intervals (CI) and hypothesis tests for a mean;
* We'll estimate the 90% CI and test the hypothesis that the population mean number of
laughs per day is different from 20... H0: μ = 20 vs. Ha: μ ≠ 20;


PROC ttest data=laughs h0=20 alpha=0.10;
      var count;
run;


/* Explanation of this PROC step:
> The DATA= option tells SAS which dataset to use; this is important if you have multiple
  datasets open.
> The h0= option specifies that the mean of the count variable should be compared to the
  null value 20 rather than the default value of 0.
> The ALPHA=0.10 option requests 90% confidence intervals; if you leave this out, 95%
  confidence intervals will be calculated by default.
> The default in SAS is to run a 2-sided test; you can change this by using the option
  SIDES=U (to test Ha: μ > 20) or SIDES=L (to test Ha: μ < 20) in the PROC TTEST line.
  This also produces a 1-sided CI, which we will not discuss this semester.
> The VAR statement indicates that the response variable is count. */
```

ANNOTATED SAS OUTPUT:

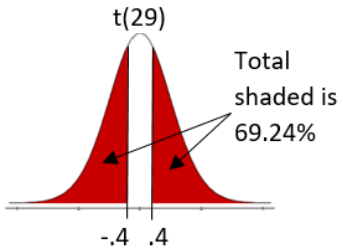


RESULTS OF THE TEST:

$H_0: \mu = 20 \; vs. \; H_a: \mu \neq 20$

From output: t = 0.40

Null distribution: t-distribution with df = 29



From output: p-value = 0.6924

- **69.24% of the null distribution is equal to or more extreme than 0.40.**

Decision: Do not reject H_0 since p-value > 0.10

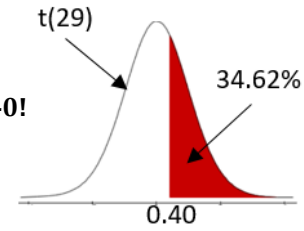
Conclusion: There is not enough evidence to say that the true mean number of laughs per day for this population is different from 20.

***WHAT IF* we had wanted to test the mean number of laughs was <u>greater</u> than 20?**

$H_0: \mu = 20 \; vs. \; H_a: \mu > 20$

Test stat stays the same: t = 0.40

Null distribution stays the same: t-distribution with df = 29



p-value now proportion of null distribution that is <u>greater</u> than or equal to 0.40!

- **Need to divide 2-sided p-value in half: $0.6924/2 = 0.3462$**
- **34.62% of the null distribution is greater than or equal to 0.40.**

Decision: Do not reject H_0 since p-value > 0.10

Conclusion: There is not enough evidence to say that the true mean number of laughs per day for this population is greater than 20.

Note: the sample mean *was* greater than 20, but not far enough above to say that the results were statistically significant! So a sample mean of 21 is consistent with the natural amount of sampling variability we would expect under this null hypothesis. Because of sampling variability, it's not enough to just look at the sample mean and make a conclusion about the population mean—sampling variability is the reason we need to conduct the hypothesis test!