

## Demo: Performing a Two-Sample t Test Using PROC TTEST

Filename: **st101d03.sas**

In this demonstration, we use PROC TTEST to perform a two-sample t test, and test whether the mean of SalePrice is the same for homes with masonry veneer as for those without.



```
PROC TTEST DATA=SAS-data-set <options>;  
  CLASS variable;  
  VAR variables < / options>;  
RUN;
```

1. Open program st101d03.sas.



```
/*st101d03.sas*/  
ods graphics;  
proc ttest data=STAT1.ameshousing3 plots(shownull)=interval;  
  class Masonry_Veneer;  
  var SalePrice;  
  format Masonry_Veneer $NoYes.;  
  title "Two-Sample t-test Comparing Masonry Veneer, No vs. Yes";  
run;  
title;
```

This PROC TTEST step doesn't use the null hypothesis option because we're testing the equality of means. The CLASS statement selects Masonry\_Veneer as the grouping variable. The CLASS statement is required in a two-sample t test. The classification variable can be numeric or character, but must have exactly two levels, because PROC TTEST divides the observations into the two groups using the values of this variable. Classification levels are determined from the formatted values of the CLASS variable, so if necessary, you can apply a format to collapse the data into two levels. The FORMAT statement here applies the \$NoYes format to display Yes and No in the output instead of Y and N.

2. Submit the code.

3. [Review the output.](#)

Start by verifying our assumption of normality of the distribution of each group by looking at the histograms and Q-Q plots. In the Summary Panel, the Distribution of SalePrice histogram appears to be normally distributed, with, of course, a different center of location. Both histograms have a blue normal reference curve superimposed on the plots to help determine whether the distributions are normal.

Consider the Q-Q plots. If the data in a Q-Q plot come from a normal distribution, the points cluster tightly around the reference line. The first plot shows homes without masonry veneer, or the Nos, and the second shows homes with, the Yess. Both Q-Q plots exhibit relatively straight lines. There's slight curvature on both, but nothing too extreme. From these four plots, it's safe to say that both populations are normally distributed.

Next, consider the Equality of Variances table. The F statistic is 1.36 and the p-value is relatively large at 0.1039. Based on this, do we reject or fail to reject the null hypothesis of equal variances? The p-value is greater than alpha, so we do not reject the null hypothesis. We don't have enough evidence to say the

variances are unequal.

Based on the results of the F test, we now look in the T-Tests table at the t test for the hypothesis of equal means, the pooled t test. The p-value is less than .0001, which is less than 0.05, so we can reject the null hypothesis that the group means are equal. We can conclude that the sale price of homes with masonry veneer is significantly different from homes without it. Also notice that the t statistic values for both the Pooled and the Satterthwaite tests are almost equal, -5.38 and -5.72. When the population variances are equal, the t values are equivalent mathematically. The slight difference here is due to random sampling differences when calculating the variances.

We can make the same conclusion about the means from the Difference Interval Plot. The Mean of SalePrice Difference (No-Yes) interval plot includes both pooled and Satterthwaite 95% confidence intervals. Notice for the pooled variance method, the confidence interval for the difference in means is between about -\$33,000 and -\$15,000. It doesn't include 0, which is our hypothesized value. In other words, we have enough evidence to say that the difference of the means is significantly different from 0 at the 95% confidence level.

Now consider the Statistics table. From our sample of 300 homes, 89 homes have masonry and 209 do not. There are 2 homes with missing data, so we'll remove these from the analysis. For the 209 homes without masonry veneer, the sale price sample mean is \$130,172, and for the 89 homes with it, the sample mean is \$154,705. The difference mean value between no masonry veneer and masonry veneer is -\$24,533. From the sample data, it's clear that homes with masonry veneer tend to have a statistically significant higher value.

---

### *Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression*

Copyright © 2019 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close