# **SASEG – Difference Testing using *t-*test**

(Spring 2016)

**Sources**(adapted with permission) **-**

Ron Freeze Course and Classroom Notes

Enterprise Systems, Sam M. Walton College of Business, University of Arkansas, Fayetteville

Microsoft Enterprise Consortium

IBM Academic Initiative

SAS® Multivariate Statistics Course Notes & Workshop, 2010

SAS® Advanced Business Analytics Course Notes & Workshop, 2010

Microsoft® Notes

Teradata® University Network

*For educational uses only - adapted from sources with permission. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission from the author/presenter.*

# **Pelican Stores - Difference Testing for Population Means**

**Example:** *Pelican Stores, a division of National Clothing, is a chain of women’s apparel stores operating throughout the country. The chain recently ran a promotion in which discount coupons were sent to customers of other National Clothing stores. Data collected for a sample of 100 in-store credit card transactions at Pelican Stores during one day while the promotion was running are contained in the file named Pelican Stores. The management is interested in knowing if the average spending by customers who are married is significantly different than the average spending by single customers.*

Let us demonstrate the use of this test statistic in the following hypothesis testing example. The population means for the total spending by categories of customers are as follows.

μ1 = the mean spending for population of customer who are married

μ2 = the mean spending for population of customer who are single

We begin with the assumption that no difference exists between the average spending by married and single customers. Hence, in terms of the mean spending, the null hypothesis is that *μ1* − *μ2* = 0. If sample evidence leads to the rejection of this hypothesis, we will conclude that the mean spending differs for the two populations. The null and alternative hypotheses for this two-tailed test are given as follows.

H0: μ1 − μ2 = 0

Ha: μ1 − μ2 ≠ 0

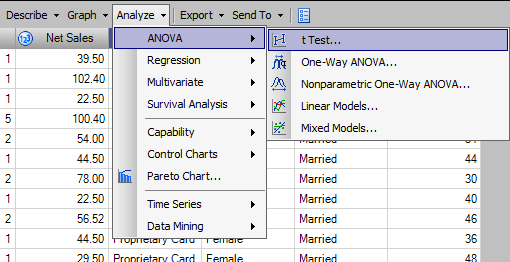
Now we will demonstrate a two-tailed hypothesis test about the difference between two population means by computing the p-value.

# C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpg**Exercise–Difference Testing for Population Means**

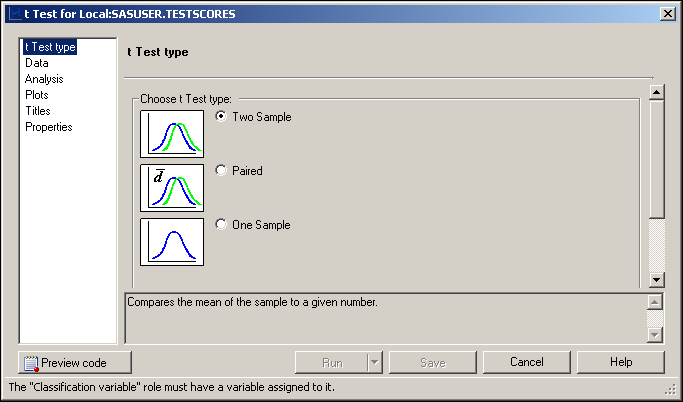
1. Open the **PelicanStores** SAS Dataset using the following path:

**File > Open >Data--> Servers >SASApp-->Files > D: > ISYS 5503--> ISYS 5503 Shared Datasets-->PelicanStores**

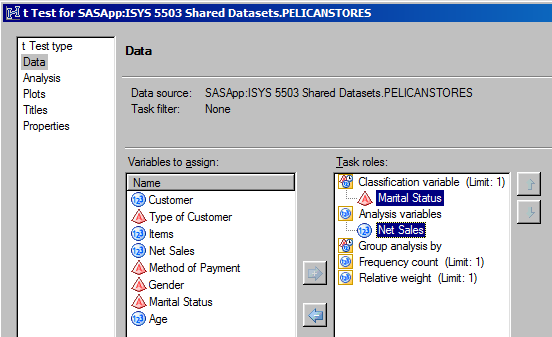
1. Select **Analze**⇨**ANOVA** ⇨t Test…



1. Leave **Two Sample** selected.



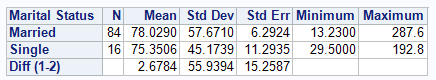
1. With **Data** selected on the left, choose Net Sales as the analysis variable and Marital Status as the classification variable.



1. Click 

**Results:**

In the following statistics table, we can see the descriptive statistics for each group and their differences. The confidence limits for the sample mean and sample standard deviation are also shown





# ***P*-value Approach**

SASEG provides both the p-value and critical value of the tests for comparison. In this section, we will provide the steps using only the p-value approach to determine whether the null hypothesis will be rejected.

### Test for Variance

Before testing the hypothesis for population mean, we will first examine the Equality of Variances table that appears at the bottom of the output. An analysis of the equality of variance tests one of the ANOVA assumptions. The result of variance test will help us determine which *t-*value to choose out of Pooled and Satterthwaite method while doing the test for mean. The hypothesis for equality of variance is given by:

H0: σ1 – σ2 = 0

Ha: σ1 – σ2 ≠ 0

Where, σ1 = Variance in spending for population of married customers, and

σ2 = Variance in spending for population of single customers

If sample evidence leads to the rejection of this hypothesis, we will conclude that the variance in spending differ for the two populations. The *F*-test result in the Equality of Variances table has a *p*-value of 0.2880. We want to be 95% confident in our conclusion, this translates to an α = 0.05. Since 0.2880 > 0.05, we will not reject the null hypothesis that the variance are equal. We conclude there is insufficient evidence to indicate that the variances are not equal.



### Test for Mean

Based on the F-test for Equality of Variances, we established that the variance for the two populations is equal. We now look at the *t*-test for the hypothesis of **Equal** means (Pooled). The null and alternative hypotheses for this two-tailed test are:

H0: μ1 − μ2 = 0

Ha: μ1 − μ2 ≠ 0

Using the Equal variance (Pooled) *t*-test, we can see that *p*-value is greater than 0.05 (Pr> |t| = 0.8610), we have insufficient evidence to reject the null hypothesis that the average spending among married and single customers are equal.



Hence, we conclude that there is no significant difference in the average spending between the population of married customers and single customers.

# **Critical Value Approach**

### Test for Variance

When we test for Equality of Variance using the critical value approach, the critical value must be looked up in an F-distribution table. In order to look up the critical value, it is necessary to know three variables: Alpha (α), Numerator Degrees of Freedom (Num DF), and Denominator Degrees of Freedom (Den DF). The format is stated as F(α, Num DF, Den DF). For α = 0.05, Num DF = 83, and Den DF = 15, the value for F(0.05,83,15) = 2.12 in the F-distribution table.

According to the rejection rule, we will reject H0 if F ≥ 2.12

However, our F-value is 1.63, which is less than 2.12; hence we will not reject the null hypothesis that the spending variances of married and single customers are equal. Therefore, we conclude that there is no significant difference in the variation of spending between the population of married customers and single customers.



### Test for Mean

We first determine the critical value and then compare it against the rejection rule. For α = 0.05 and Degrees of Freedom or DF = 98, the corresponding value for t.05 = 1.661 in the *t-*distribution table.

According to the rejection rule, we will reject H0 if t ≥ 1.661

However, our t-value is 0.18, which is less than 1.661, hence we will not reject the null hypothesis that the average spending among married and single customers are equal. Therefore, we conclude that there is no significant difference in the average spending by population of married customers and single customers.

# C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpg**Additional Questions**

1. Using the PelicanStore SAS Dataset, test the hypothesis that the average spending is equal for different Gender (Male and Female). Examine the data using the Summary Statistics table. What information can you obtain from looking at the data?
2. Test the hypothesis that the average spending is equal for different Types of Customer (Regular and Promotional).