# **SASEG – Hypothesis Testing for Proportions**

(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.*

# **Eagle Outfitters – Hypothesis Testing for Proportion**

**Example:** *Eagle Outfitters is a chain of stores specializing in outdoor apparel and camping gear. They are considering a promotion that involves mailing discount coupons to all their credit card customers. This promotion will be considered a success if more than 10% of those receiving the coupons use them. Before going national with the promotion, coupons were sent to a sample of 100 credit card customers.*

In order to know whether the promotion will be a success, we develop the following two hypotheses that can be used to test whether the population proportion of those who will use the coupons is sufficient to go national.

1. **Use case: Coupons used = Yes**

*H*0: *pyes ≤* 0.10

*H*a: *pyes* > 0.10

1. **Use case: Coupons used = No**

*H*0: *pno ≥* 0.90

*H*a: *pno* < 0.90

When running the proportional analysis, SASEG tests the variable that is first in ascending order. In this case, SASEG would test the hypothesis where the variable ‘Coupons used’ = **No**. Hence, we will use the second use case (Coupons used = No) in this illustration. To determine if the proportion of population using discount coupons is greater than 10%, an upper tail test with *Ha: pno*< 0.90 is appropriate. The null and alternative hypotheses for the test would be:

*H*0: *pno ≥* 0.90

*H*a: *pno*< 0.90

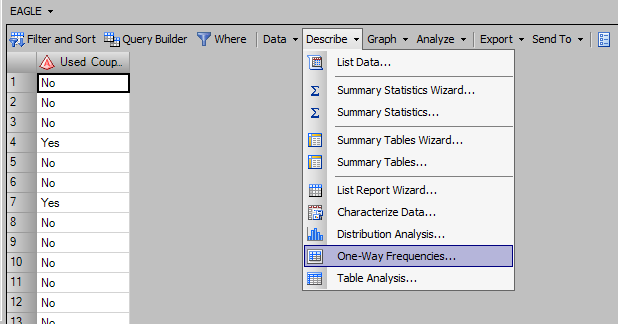
If H0 can be rejected, the test results will lend statistical support for the conclusion that the population proportion using coupons is greater than 10% (or population proportion not using coupons is less than 90%) and sufficient to go national. We will use a 95% significance level (α= 0.05) for this hypothesis test.

## C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise 1 –Proportions of Coupon Use

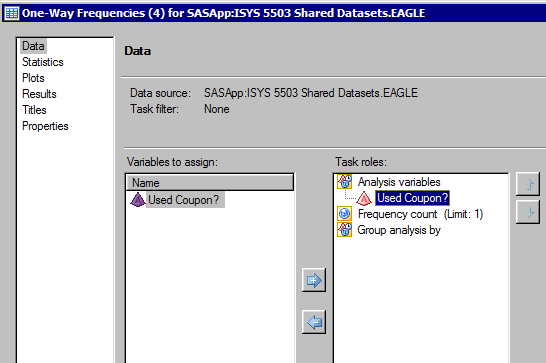
1. Open the **Eagle** SAS Dataset using the following path:

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

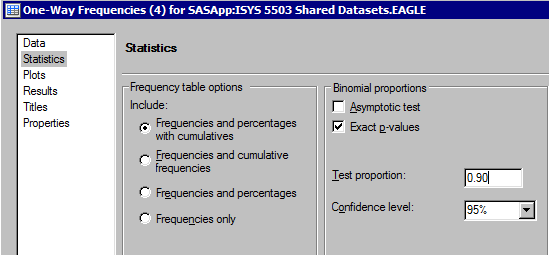
1. Select **Describe**⇨**One-Way Frequencies**



1. With **Data** selected on the left, drag and drop **Used Coupon** variable to the Analysis Variables field

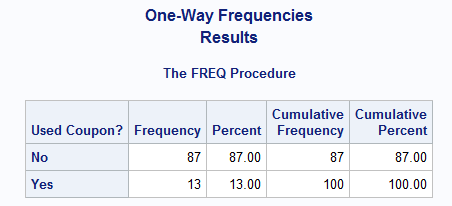
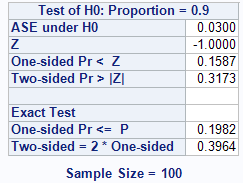


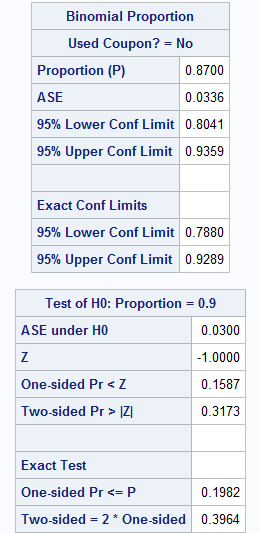
1. With **Statistics** selected on the left, check **Exact p-values** box in the Binomial proportions section. Change **Test proportion** to 0.90. NOTE: the confidence level default was 95%.



5. Click 

**Results:**





As explained earlier in the exercise, SASEG tested the hypothesis for Coupons Used = **No**

In the table Test of H0: Proportion = 0.9, the Exact Test use is the Oned-sided Pr <= P. Because the *p*-value is greater than α = 0.05 (Pr ≤ P = 0.1982),we have insufficient evidence to reject the null hypothesis that the population proportion using coupons is greater than 10% (or population proportion not using coupons is less than 90%). Therefore, our hypothesis test presents strong evidence to not go national with the campaign.

# **Pelican Stores – Hypothesis Testing for Proportion**

**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 as ample 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. Since the promotion is related to married customers, the campaign will be considered a success if more than 75% of those using the coupons is married. If the promotion is successful, Pelican Stores will launch the campaign nationally.*

In order to know whether the promotion will be a success if launched nationally, we can develop the following hypothesis that can be used to test whether the population proportion of married customers using the coupons is greater than 75%. NOTE: Either hypothesis test is correct for answering the Pelican Stores question.

1. **Use case: Marital Status = Married**

*H*0: *pmarried ≤* 0.75

*H*a: *pmarried* > 0.75

1. **Use case: Marital Status = Single**

*H*0*: psingle ≥ 0.25*

*H*a*: psingle < 0.25*

When running the proportional analysis, SASEG tests the variable that is first in ascending order. In this case, SASEG would test the hypothesis where the variable ‘Marital Status’ = **Married**. Hence, SASEG will test the first use case (Marital Status = Married) in this illustration. To determine whether the proportion of married customers using discount coupons is greater than 75%, an upper tail test with *Ha: pmarried* > 0.75 is appropriate. The null and alternative hypotheses for the test would be:

*H*0: *pmarried ≤* 0.75

*H*a: *pmarried* > 0.75

If H0 can be rejected, the test results will lend statistical support for the conclusion that the population proportion of married customers using discount coupons is greater than 75%. We will use a 95% significance level (α= 0.05) for this hypothesis test.

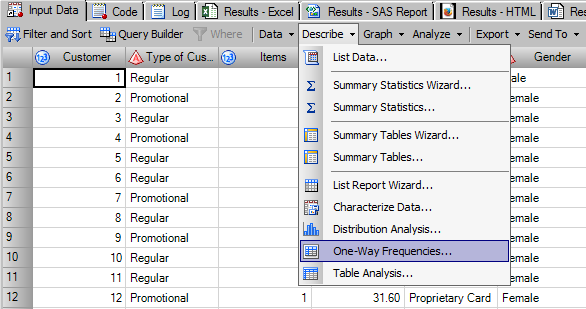
NOTE: we will also need to filter our data to include only those customers using the promotional coupons.

## C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise 2 –Proportions for Married Coupon Use

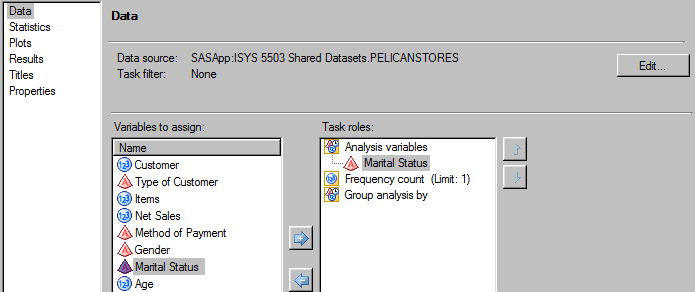
1. Open the **PelicanStores** SAS Dataset using the following path:

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

1. Select **Describe**⇨**One-Way Frequencies**

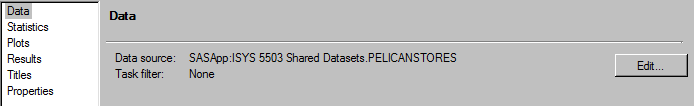


1. With **Data** selected on the left, drag and drop **Marital Status** variable to the Analysis Variables field

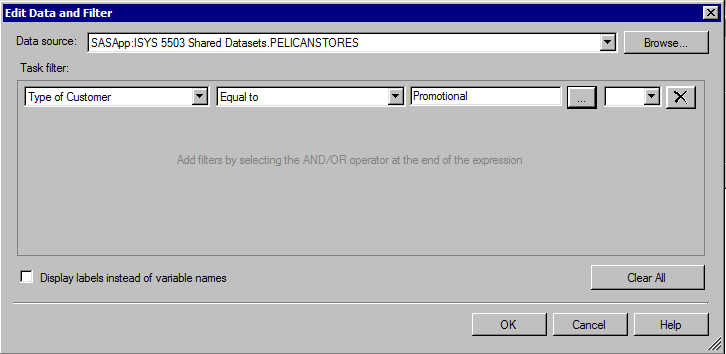


Since our interest is in the proportion of the customers **using the coupons (or Promotional)** that are married, we need to filter the data set to only those customers using the coupons.

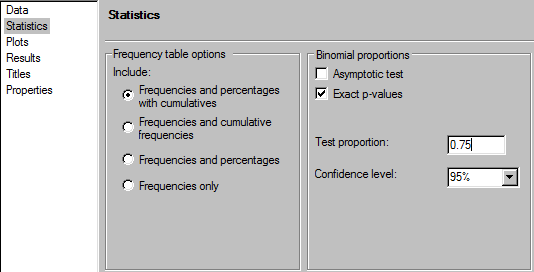
1. On the **Data** screen, select **Edit…** on the right.



1. Under the task filter, select “Type of Customer”, “Equal to”, and “Promotional” as shown in the screenshot below. Click OK.

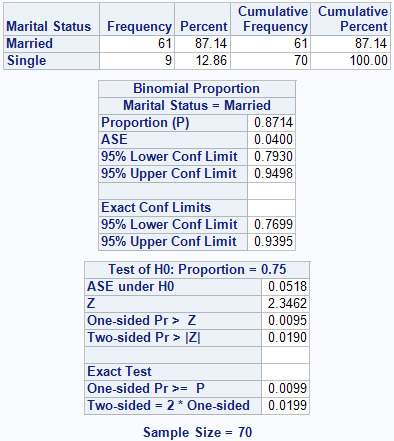


1. With **Statistics** selected on the left, check **Exact p-values** box in the Binomial proportions section. Change **Test proportion** to 0.75



1. Click 

**Results:**



Note the sample size to be 70, as the other 30 regular customers (who did not use promotional coupons) were filtered out

As explained earlier in the exercise, SASEG tested the hypothesis for Marital Status = **Married**

The test results indicate that the *p*-value is less than 0.05 (Pr ≤ P = 0.0099), hence we have strong evidence to reject the null hypothesis and accept that the alternative hypothesis (*H*a) is true. The alternative hypothesis states that the population proportion using the coupons and married is greater than 75%. Therefore, we conclude that the promotion will be successful if launched nationally.

# **Pelican Stores – Hypothesis Testing for Standard Deviation**

Variance is a measure that is useful in comparing the amount of variability of two or more variables. The standard deviation is defined to be the positive square root of the variance. The sample standard deviation is denoted by ‘s’ and population standard deviation is denoted by ‘σ*’.*

In Pelican Stores case, it is of particular interest to know the standard deviation of customers across both the categories, single and married. Customers with larger standard deviation (more variability) carry a larger risk as their behavior is uncertain and cannot be predicted accurately. It would be interesting for the seller to know which segment of customers has higher or lower standard deviation and if there is any significant difference between the segment of customers in terms of variance in their spending. For example, it may be assumed that due to lifestyle differences, married customers have a set pattern of buying behavior with little or no change while customers who are single may have a higher degree of randomness in buying behavior.

In order to test this variability, we first must look at the descriptive statistics for the sample data set.

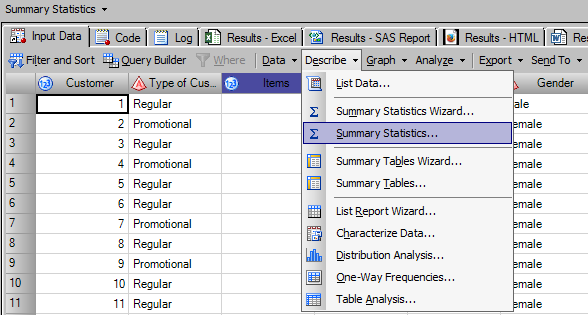
NOTE: this hypothesis test assumes that the population standard deviation is not known.

In SASEG, we can find the standard deviation for different variables with the help of descriptive statistics. Descriptive statistics provide simple summaries about the sample and the measures. The following exercise focuses on running descriptive statistics for the variable ‘Age’ within different marital groups.

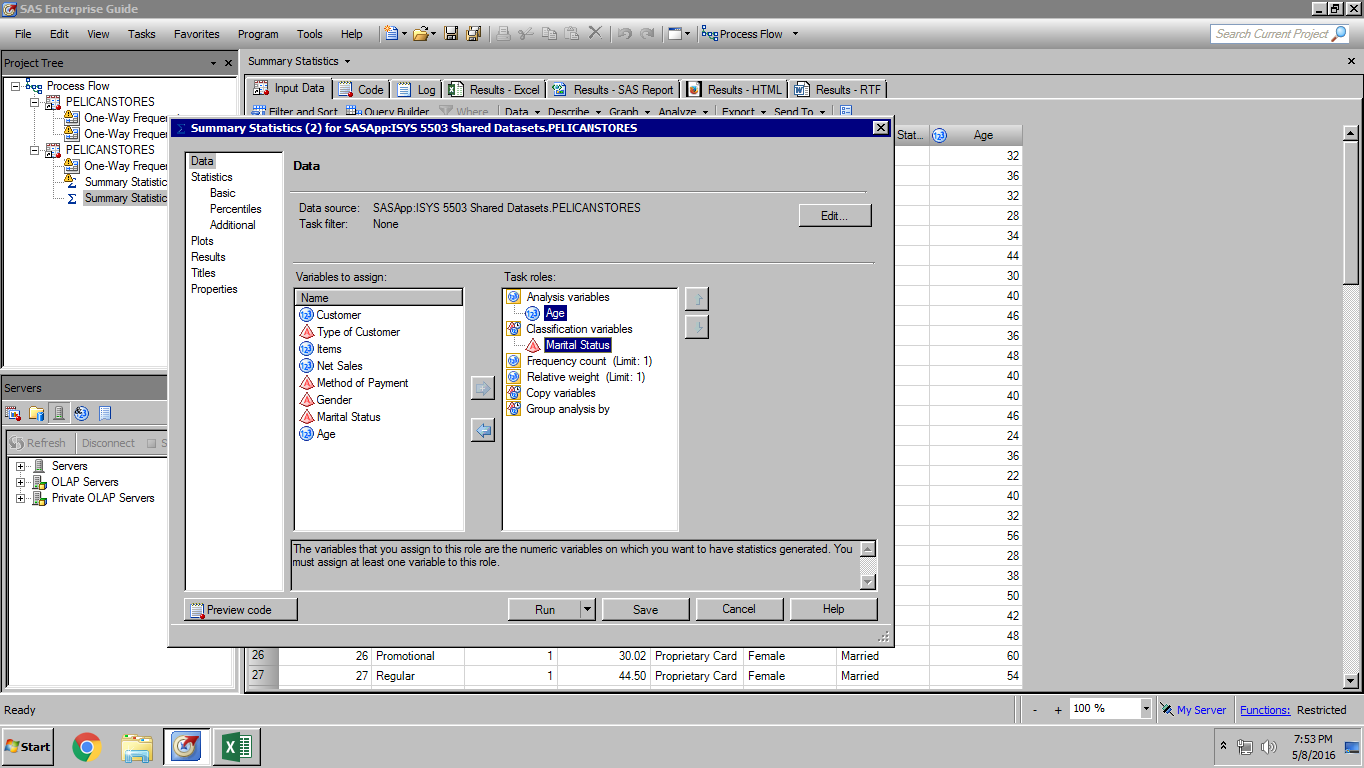
## C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise 3 – Descriptive Statistics

1. Click **Input Data** and select **Describe**⇨**Summary Statistics…** from the drop-down menus.

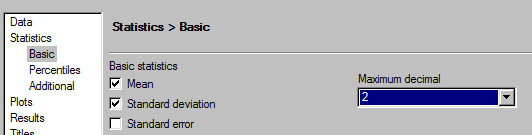
**🖉** If you close the data table first, then you will have to click **Tasks**⇨**Describe**⇨  
**Summary Statistics…** from the top menu bar.



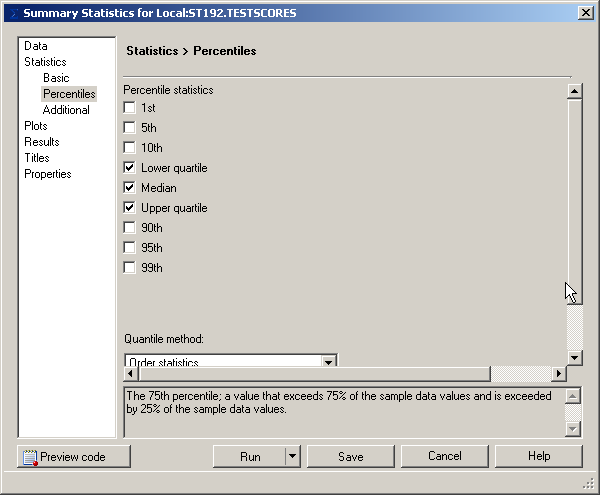
1. With **Data** selected on the left, drag and drop **Age** to Analysis variable and **Marital Status** to Classification variable



1. Select **Basic** under Statistics on the left. Leave the default basic statistics. Change Maximum decimal places to **2**.

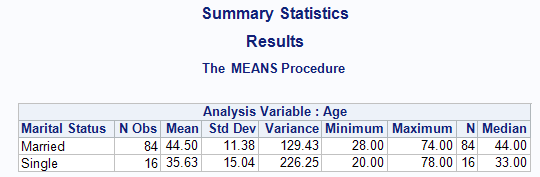


1. Select **Percentiles** on the left. Under Percentile statistics, check the boxes for **Lower quartile**,**Median**, and **Upper quartile**.

****

1. Select **** to run the analysis.

The result is shown below:



We can see that the Standard deviation in the age of Married and Single customers is 11.38 and 15.04 respectively.

## C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise 4 – Hypothesis testing for Standard Deviation

Suppose we want to know if the variation in age of Married customers is equal to that of Single customers. In order to test the difference of standard deviation of age of Married and Single customers, we will form the following hypothesis:

*H*0: *σ1 = σ2*

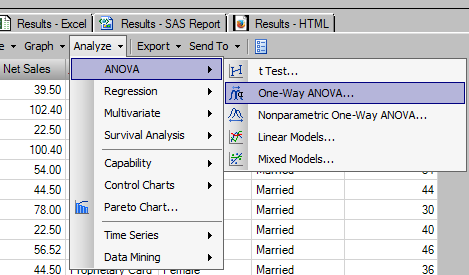
*H*a: *σ1 ≠ σ2*

Where, *σ1* = Standard deviation of age of Married customers and *σ2* = Standard deviation of age of Single customers.

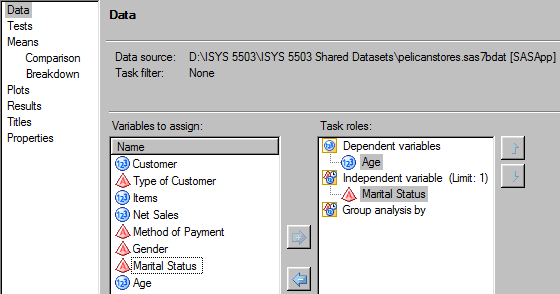
If the larger variance (or standard deviation) is present in the first sample, then the test is right-tailed. Otherwise, the test is left-tailed. We will perform the test at a significance level of α = 0.05.

1. Click **Input Data** and select **Analyze**⇨**ANOVA**⇨**One-Way ANOVA…** from the drop-down menus.

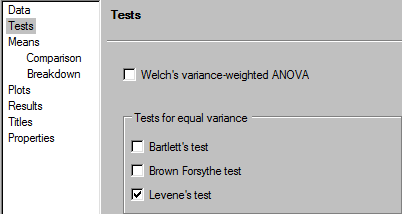
**🖉** If you close the data table first, then you will have to click **Tasks**⇨**ANOVA**⇨**One-Way ANOVA…** from the top menu bar.



1. With **Data** selected on the left, drag and drop **Age** to Dependent Variable and **Marital Status** to Independent Variable



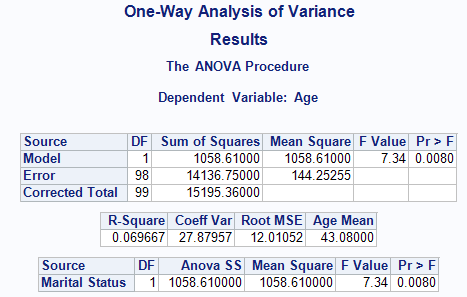
1. With **Tests** selected on the left, check the box against ‘Levene’s Test’.



1. Click**** to run the analysis

**Test Result:**

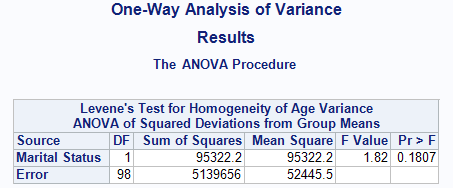
The SASEG provides F-test results for variance (Levene’s test) as well as for mean.



Hypothesis test for difference in mean age for both categories of customer population

The hypothesis test for mean provides a *p-*value of 0.0080. At a significance level of α = 0.05, the *p*-value (0.0080) is less than 0.05, which indicates an overwhelming evidence to reject the null hypothesis that mean age of married and single customer population are equal. We conclude that there is a significant difference in the mean age of single and married customer population.

Test for variance:



Hypothesis test for difference in variance of age for both categories of customer population

The result of the F-test for equal variance (Levene’s test) provides a *p-*value of 0.1807. Since our selected confidence level is 95% and 0.1807 >α = 0.05, we have insufficient evidence to reject the null hypothesis that there is no significant difference in the age variance for single and married customer population.

This finding leads to the conclusion that the two groups of customers, single and married, do not differ in terms of age variance.

Additional Question

1. Using the PelicanStore SAS Dataset, test the hypothesis that variation in spending by married customers is equal to variation in spending by single customers.