Two-Sample Tests for Means

Dr Tom Ilvento
Department of Food and Resource Economics
University of Delaware

Overview

- We have made an inference from a single sample mean and proportion to a population, using
  - The sample mean (or proportion)
  - The sample standard deviation
  - Knowledge of the sampling distribution for the mean (proportion)
  - And it matters if sigma is known or unknown; large or small sample; whether the population is distributed normally
- The same strategy will apply for testing differences between two means or proportions
- These will be two independent, random samples
- We will have
  - Sample Estimate (of the difference)
  - Hypothesized value
  - Standard error
  - Knowledge of a sampling distribution

Testing differences between two means or proportions

- We can make a point estimate and a hypothesis of the difference of the two means
- Or a Confidence Interval around the difference of the two means
- With a few twists
- Mean
  - Sigma known or unknown
  - Should we pool the variance?
- Proportions
  - When testing Ho: we need to check if $p_1 = p_2$

Testing differences between two means or proportions

- We will also need to come up with:
  - An **estimator** of the difference of two means/ proportions
  - The **standard error** of the sampling distribution for our estimator
- With two sample problems we have **two sources of variability** and the sampling error must take that into account
- We also must assume the samples are **independent random samples**

The Estimator

- How do I make a comparison of two means or proportions?
  - **Ratio of the two** - if they are equal we will have a ratio near 1.0
  - **Difference** - if they are equal we will have a difference near zero
- A **Difference** is preferred in this case
  - For the population $\mu_1 - \mu_2 = 0$
  - From the sample $mean_1 - mean_2 = 0$

The Standard Error

- For a single mean the standard error looks like this

$$\sigma_{(\bar{x}_1)} = \sqrt{\frac{\sigma_1^2}{n_1}}$$

- Now we are looking at two independent random samples
- And the standard error will look something like this

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

What do we mean by Independent Random Samples?

- **Independent samples** means that each sample and the resulting variables do not influence the other sample
  - If we sampled the same subjects at two different times we would not have independent samples
  - If we sampled husband and wife, they would not be independent – their responses should be related to each other!
- However, we have a strategy to assess non-independent samples – paired difference test

Decision Table for Two Means

- Use this table to help in the Difference of Means Test
- **Small sample problems require us to assume normal distributions, and we <u>should</u> pool if possible**

Targets	Assumptions	Test Statistic
H_0: $\mu_1 - \mu_2 = D$	Independent Random Samples, **Sigma Known**	Use σ_1 and σ_2; and **standard normal** for comparisons
	Independent Random Samples, **Sigma Unknown**	Use s_1 and s_2; and **t-distribution** for comparisons
	Independent Random Samples, Sigma Unknown; we can assume **variances are equal**	Use t-distribution Use a single estimate of the variance, called a ***"Pooled Variance"***

Decision Table for Two Proportions

- Use this table to help in the Difference of Proportions Test
- **Large sample problems require that the combination of p*n or q*n > 5**

Targets	Assumptions	Test Statistic
$H_0: p_1 - p_2 = D$	Independent Random Samples; **Large sample sizes (n1 and n2 >50 when p or q > .10); Ho: D = 0**	**standard normal** for comparisons with z; **Pool the estimate of P** based on the Null Hypothesis
	ndependent Random Samples; **Large sample sizes (n1 and n2 >50 when p or q > .10); Ho: D≠ 0**	**standard normal** for comparisons with z;

Example of Difference of Mean: Weight of Pallets of Roof Shingles

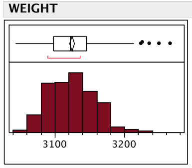
- Studies have shown the weight is an important customer perception of quality, as well as a company cost consideration.
- The last stage of the assembly line packages the shingles before placement on wooden pallets.
- The company collected data on the weight (in pounds) of pallets of Boston and Vermont variety of shingles.
- You can try this yourself: Pallet.xls or Pallet.jmp

Example of Difference of Mean: Weight of Pallets of Roof Shingles

- **What do you see?**
- **For Boston**
  - **Mean = 3124.215**
  - **Median = 3122**
  - **Std Dev = 34.713**
- **For Vermont**
  - **Mean = 3704.042**
  - **Median = 3704**
  - **Std Dev = 46.744**
- Both look approximately normal
- The variances are very similar

Distributions TYPE=Boston

WEIGHT



Quantiles

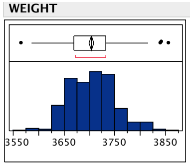
100.0%	maximum	3266.0
99.5%		3252.5
97.5%		3201.6
90.0%		3166.0
75.0%	quartile	3146.0
50.0%	median	3122.0
25.0%	quartile	3098.0
10.0%		3080.0
2.5%		3064.0
0.5%		3045.7
0.0%	minimum	3044.0

Moments

Mean	3124.215
Std Dev	34.713
Std Err Mean	1.810
upper 95% Mean	3127.773
lower 95% Mean	3120.656
N	368.000
Sum Wgt	368.000
Sum	1149711.0
Variance	1204.992
Skewness	0.525
Kurtosis	0.773
CV	1.111
N Missing	0.000

Distributions TYPE=Vermont

WEIGHT



Quantiles

100.0%	maximum	3856.0
99.5%		3846.8
97.5%		3804.0
90.0%		3763.8
75.0%	quartile	3732.0
50.0%	median	3704.0
25.0%	quartile	3670.0
10.0%		3646.2
2.5%		3626.0
0.5%		3579.1
0.0%	minimum	3566.0

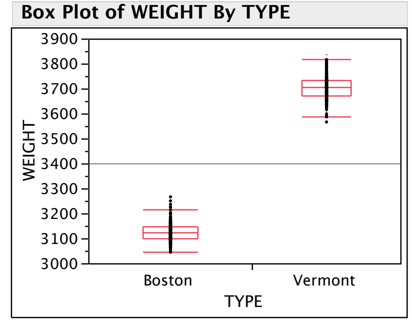
Moments

Mean	3704.042
Std Dev	46.744
Std Err Mean	2.573
upper 95% Mean	3709.104
lower 95% Mean	3698.980
N	330.000
Sum Wgt	330.000
Sum	1222334.0
Variance	2185.032
Skewness	0.287
Kurtosis	0.212
CV	1.262
N Missing	0.000

Example of Difference of Mean: Weight of Pallets of Roof Shingles

- We can clearly see that the mean weight of Vermont shingle pallets is higher
- And the spread of the two types, the variance or standard deviation are not much different
  - The means are different
  - But the spread is not

Box Plot of WEIGHT By TYPE



Example of Difference of Mean: Weight of Pallets of Roof Shingles

- **Ho:**
- **Ha:**
- **Assumptions**
- **Test Statistic**
- **Rejection Region**
- **Calculation:**
- **Conclusion:**

- **Ho: $\mu_v - \mu_b$ = ?**
- **Ha: $\mu_v - \mu_b \neq$? 2-tailed test**
- **large samples; sigma unknown; could pool**
- **t* = ?**
- **α = .05, $t_{\alpha/2,\, d.f.}$ = ?**
- **t* = ?**

We need to think of the sampling distribution of the difference of two means

- The mean of the sampling distribution for ($mean_1$-$mean_2$)
- The sampling distribution center will equal ($\mu_1 - \mu_2$)
- The difference is hypothesized = Do
  - We usually designate the expected difference as Do under the the null hypothesis
  - Most often we think of Do = 0; no difference
  - **But it could be something else**

The Standard Error of the Difference of Two Means

- The Standard Error of the sampling distribution difference of two means is given as:
- The sampling distribution of ($mean_1$-$mean_2$) is approximately normal for large samples under the Central Limit Theorem
- It is based on two independent random samples
- We typically use the sample estimates of s_1 and s_2
- **And then use the t-distribution for the test**

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The Test Statistic

- The test statistic involves the difference of two means
  - Compared to a difference specified by the Null Hypothesis
  - Noted as Do
  - Divided by the standard error
- The calculation for our example is:

$$t^* = \frac{(3124.2 - 3704.0) - 0}{}$$

$$SE = \sqrt{\frac{1204.99}{368} + \frac{2185.03}{330}} = \sqrt{3.2744 + 6.6213} = 3.1457$$

t* = (-579.8 – 0)/3.15 = -184.31

Example of Difference of Mean: Weight of Pallets of Roof Shingles

- Ho:
- Ha:
- Assumptions
- Test Statistic
- Rejection Region
- Calculation:
- Conclusion:

- Ho: $\mu_B - \mu_V = 0$
- Ha: $\mu_B - \mu_V \neq 0$ 2-tailed test
- large samples; sigma unknown; could pool
- t* = (-579.8 – 0)/3.15
- $\alpha = .05$, $t_{.05/2,\ 603\ d.f.} = -1.964$ or 1.964
- t* = -184.31
- This value is HUGE!!
- Reject Ho: $\mu_B - \mu_V = 0$
- There is a difference!

What is the p-value for our test statistic?

- t* = -184
- IT IS HUGE!!!!
- The p-value is smaller than = .05/2
  - p < .001
- Therefore, **we reject** Ho: $\mu_B - \mu_V = 0$

To do this test with Excel

- **The data needs to be in two columns - one for each group**
- **Data Analysis**
  - t-test: two sample assuming unequal variances
  - Pick both variables
  - Note labels or not
  - Specify the difference under a Null Hypothesis
  - Tell Excel where to put the output
- Dress it up

Excel Results

- Look to see that you can find all the information
  - The means
  - The variances
  - The Hypothesized mean difference
  - df
  - The Test Statistic
- Excel gives the critical values and p-values for both a one and two-tailed test

t-Test: Two-Sample Assuming Unequal Variances

	Boston	*Vermont*
Mean	3124.215	3704.042
Variance	1204.992	2185.032
Observations	368	330
Hypothesized Mean Difference	0	
df	603	
t Stat	-184.321	
P(T<=t) one-tail	0	
t Critical one-tail	1.647385	
P(T<=t) two-tail	0	
t Critical two-tail	1.963906	

$$d.f. = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\left(\frac{\left(s_1^2/n_1\right)^2}{(n_1-1)} + \frac{\left(s_2^2/n_2\right)^2}{(n_2-1)}\right)}$$

JMP Results

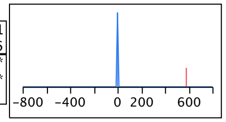
- Most of the same information is here from JMP
- Plus it gives the **confidence interval**

t Test

Vertmont-Boston

Assuming unequal variances

Difference	579.828	t Ratio	184.321
Std Err Dif	3.146	DF	602.7216
Upper CL Dif	586.006	Prob > |t|	0.0000*
Lower CL Dif	573.650	Prob > t	0.0000*
Confidence	0.95	Prob < t	1.0000



$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,df} s_{(\bar{x}_1 - \bar{x}_2)}$$

$$(3704.04 - 3124.22) \pm 1.964(3.146) = 579.82 \pm 6.18$$

573.65 to 586.01

What if we were to assume the variances were equal?

- First, you can just assume it - it has to be reasonable
  - A ratio of the two variances would be the way to test it
  - The ratio should be about 1
  - with some sampling error
- Our ratio is **2185.032/1204.992 = 1.81**

If we can assume the variances are equal

- There may be times when we think the difference between the two samples is primarily the means
- But the variances are similar
- In this case we ought to use information from both samples to estimate sigmas
- We will use a **t-test** and the t distribution and adjust the degrees of freedom
- Assumptions
  - The **population variances are equal**
  - **Random** samples selected **independent**ly of each other
- **Pooling the variances is critical for small sample difference of means problems!**

Pooling the Variances

- If we can assume ($s_1 = s_2$), we should use information from both sample estimates
- **First Step:** calculate pooled variance using information from both samples
- **Step 2:** Use the pooled estimate of the variance to calculate the standard error

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Note: the denominator reduces to ($n_1 + n_2 - 2$) which is the d.f. for the t distribution

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

What does Pooling do for us?

- Pooling generates a weighted average as the estimate of the variance
- The weights are the sample sizes for each sample
- A pooled estimate is thought to be a better estimate if we can assume the variances are equal
- And our degrees of freedom are larger - d.f. = n1 + n2 – 2
- Which means the t-value will be smaller
- **Note: if $n_1 = n_2$, the formula simplifies to $(s^2_1+s^2_2)/2)$**

Excel Assuming Equal Variances

t-Test: Two-Sample Assuming Equal Variances

	Boston	*Vermont*
Mean	3124.215	3704.042
Variance	1204.992	2185.032
Observations	368	330
Pooled Variance	1668.258	
Hypothesized Mean Difference	0	
df	696	
t Stat	-187.2495	
P(T<=t) one-tail	0	
t Critical one-tail	1.647046	
P(T<=t) two-tail	0	
t Critical two-tail	1.963378	

The d.f. increased

The test statistic changed slightly because the standard error changed when we used a pooled estimate of s^2

Equal Variances with JMP

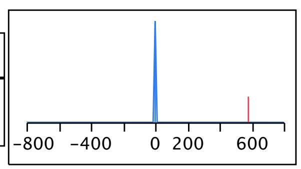
- We get the same results with JMP

t Test

Vermont-Boston

Assuming equal variances

Difference	579.828	t Ratio	187.2495
Std Err Dif	3.097	DF	696
Upper CL Dif	585.907	Prob > |t|	0.0000*
Lower CL Dif	573.748	Prob > t	0.0000*
Confidence	0.95	Prob < t	1.0000



Small Sample Problem

- Federal regulations require that certain materials, such as children's pajamas, be treated with a flame retardant.
- An evaluation of a flame retardant was conducted at two different laboratories. While there may be measurement error associated with the lab work, we should not expect systematic differences between two laboratories.
- An experiment was designed so that each laboratory received the same number of samples of three different materials - 9 samples per laboratory
- The data are the length of the charred portion of the material.
- Test to see if the there is a difference in the measurements between the two laboratories at α = .01.

Decision Table for Two Means

- Use this table to help in the Difference of Means Test
- **Small sample problems require us to assume normal distributions, and we <u>should</u> pool if possible**

Targets	Assumptions	Test Statistic
H_0: $\mu_1 - \mu_2 = D$	Independent Random Samples, **Sigma Known**	Use σ_1 and σ_2; and **standard normal** for comparisons
	Independent Random Samples, **Sigma Unknown**	Use s_1 and s_2; and **t-distribution** for comparisons
	Independent Random Samples, Sigma Unknown; we can assume **variances are equal**	Use t-distribution Use a single estimate of the variance, called a ***"Pooled Variance"***

Output from JMP

RATING From Both Labs

Quantiles

100.0%	maximum	4.3000
99.5%		4.3000
97.5%		4.3000
90.0%		4.1200
75.0%	quartile	3.5250
50.0%	median	2.9500
25.0%	quartile	2.4500
10.0%		2.1700
2.5%		1.9000
0.5%		1.9000
0.0%	minimum	1.9000

Moments

Mean	3.0055556
Std Dev	0.6829626
Std Err Mean	0.1609758
upper 95% Mean	3.3451849
lower 95% Mean	2.6659262
N	18
Sum Wgt	18
Sum	54.1
Variance	0.4664379
Skewness	0.3578967
Kurtosis	-0.684997
CV	22.72334
N Missing	0

Stem and Leaf

Stem	Leaf	Count
4	13	2
3	569	3
3	1123	4
2	56778	5
2	233	3
1	9	1

1|9 represents 1.9

RATING Lab 1

Quantiles

100.0%	maximum	4.3000
99.5%		4.3000
97.5%		4.3000
90.0%		4.3000
75.0%	quartile	4.0000
50.0%	median	3.5000
25.0%	quartile	3.1500
10.0%		2.8000
2.5%		2.8000
0.5%		2.8000
0.0%	minimum	2.8000

Moments

Mean	3.533
Std Dev	0.492
Std Err Mean	0.164
upper 95% Mean	3.912
lower 95% Mean	3.155
N	9.000
Sum Wgt	9.000
Sum	31.800
Variance	0.242
Skewness	0.211
Kurtosis	-0.898
CV	13.937
N Missing	0.000

RATING Lab 2

Quantiles

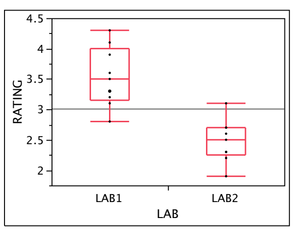
100.0%	maximum	3.1000
99.5%		3.1000
97.5%		3.1000
90.0%		3.1000
75.0%	quartile	2.7000
50.0%	median	2.5000
25.0%	quartile	2.2500
10.0%		1.9000
2.5%		1.9000
0.5%		1.9000
0.0%	minimum	1.9000

Moments

Mean	2.478
Std Dev	0.349
Std Err Mean	0.116
upper 95% Mean	2.746
lower 95% Mean	2.209
N	9.000
Sum Wgt	9.000
Sum	22.300
Variance	0.122
Skewness	0.148
Kurtosis	0.371
CV	14.093
N Missing	0.000

JMP Difference of Means Test assuming equal variances

- Is this a one or two-tailed test? **Ho: $\mu_1 - \mu_2 = 0$ Ha: $\mu_1 - \mu_2 \neq 0$**
- What is the Pooled estimate of the variance? **(.242+.122)/2 = .182**
- What is the Standard Error? **SQRT(.182/9+.182/9) = .201**
- How many degrees of freedom for our test? **9 + 9 - 2 = 16**
- What is your conclusion? **p-value is < .0001; Reject Ho**

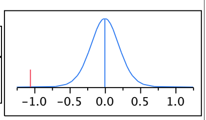


t Test

LAB2-LAB1

Assuming equal variances

Difference	-1.0556	t Ratio	-5.2455
Std Err Dif	0.2012	DF	16
Upper CL Dif	-0.6290	Prob > |t|	<.0001*
Lower CL Dif	-1.4821	Prob > t	1.0000
Confidence	0.95	Prob < t	<.0001*



Summary

- The difference of means hypothesis test follows a similar format as a single mean or proportion hypothesis:
  - Sample estimate; Standard error; Null and alternative hypotheses
  - Set an alpha level or use a p-value
- The Confidence Interval will be similar as well
- For hypotheses tests, we will be asked if we feel the variances are equal or not – we will pool the variances if yes
- For small sample difference of means problems, when n_1 and n_2 are less than 30,
  - we must be able to assume the variables are distributed approximately normal
  - and we would like to assume the variances are equal