# Tests and Confidence Intervals for Two Means

**Read: Sections 2.7 and 2.8 of *Dielman***

- Do advertisements help to increase store sales?
- Data from two independent samples
  - Analysis assuming equal variances
  - Analysis allowing variances to be different
- From paired samples

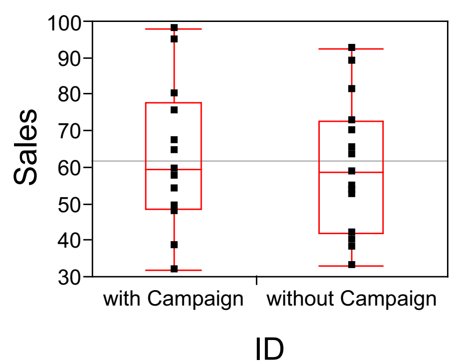


## Example: The Effect of an Ad Campaign on Store Sales

A national chain of clothing stores wishes to investigate the effect of an intensive in-store ad campaign on store sales.

They begin with a RANDOM sample of 28 stores.
In 13 of these stores they run the ad campaign. In the remaining 15 they do not.

Here are side-by-side boxplots for the weekly sales (in $1,000) in these stores.



## Summary Statistics from JMP

| Sample Measure | With Campaign | Without Campaign |
|---|---|---|
| Mean | $\bar{Y}_1 = 62.85$ | $\bar{Y}_2 = 60.35$ |
| Std Dev | $s_1 = 20.03$ | $s_2 = 18.39$ |
| $n$ | $n_1 = 13$ | $n_2 = 15$ |
| Std Error Mean | 5.55 | 4.75 |
| Upper 95% Mean | 75.0 | 70.5 |
| Lower 95% Mean | 50.7 | 50.2 |

Formulas: $\bar{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}$, etc. and $s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(Y_{1i} - \bar{Y}_1\right)^2$

Need to use sample means $\bar{Y}_1$ and $\bar{Y}_2$ to test if the two population means are equal- *ie*, if

$$\mu_1 = \mu_2$$

Notice the two population standard deviations $\sigma_1$ and $\sigma_2$ are unknown too.



## Basic Statistical Setting:

- Two random samples
  - Populations assumed to be normal:
    - With population means $\mu_1$ and $\mu_2$
    - With population standard deviations $\sigma_1$ and $\sigma_2$
    - *Independent* samples with sample sizes $n_1$ and $n_2$
  - Statistics computed from the samples:
    - Sample means $\bar{Y}_1$ and $\bar{Y}_2$
    - Sample standard deviations $s_1$ and $s_2$

- Goal = comparisons of the two population means - primarily
  - a. Tests of $H_0: \mu_1 = \mu_2$ [or of $H_0: \mu_1 \le \mu_2$ or of $H_0: \mu_1 \ge \mu_2$] or
  - b. Confidence intervals for the difference $\mu_1 - \mu_2$

**Fact**: $\overline{Y}_1 - \overline{Y}_2$ is a good estimator of $\mu_1 - \mu_2$.

We also need the standard deviation of $\overline{Y}_1 - \overline{Y}_2$. This is

$$SD\left(\overline{Y}_1 - \overline{Y}_2\right) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \ .$$

The estimate of this is called the **Standard Error**.

- Analysis assuming equal variances:

If we assume $\sigma_1^2 = \sigma_2^2 \triangleq \sigma_p^2$ (say) then we can estimate $\sigma_p^2$ from the data by $s_p^2$, say –

*See next page for formula for* $s_p^2$.

The SE is then

$$SE_p = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} .$$

- Analysis not assuming equal variances: Then we need separate estimates $s_1^2$ and $s_2^2$ for $\sigma_1^2$ and $\sigma_2^2$; and the SE is

$$SE_{unp} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} .$$



## Different classes of procedures

**A.** Assume $\sigma_1 = \sigma_2$

Use "pooled variance" procedures:

Estimate the common variance $[\sigma_1^2 = \sigma_2^2]$ as

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{1}{n_1 + n_2 - 2}\left(\sum_{i=1}^{n_1}\left(Y_{1i} - \overline{Y}_1\right)^2 + \sum_{i=1}^{n_2}\left(Y_{2i} - \overline{Y}_2\right)^2\right).$$

Estimate the S.E. of $\overline{Y}_1 - \overline{Y}_2$ as

$$SE_p = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \ .$$

**B.** Do not assume $\sigma_1 = \sigma_2$

Don't "pool"; and estimate the S.E. of $\overline{Y}_1 - \overline{Y}_2$ as

$$SE_{unp} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} .$$

**C.** Assume the data have a matched-pairs structure, & use matched-pairs method.

## Confidence Intervals for $\mu_1 - \mu_2$

**A.** Assume $\sigma_1 = \sigma_2$

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{\alpha/2;\, n_1+n_2-2} SE_p$$

**B.** Don't assume $\sigma_1 = \sigma_2$

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{\alpha/2;\, \Delta} SE_{unp};$$

See p. 41 for definition of $\Delta$. JMP provides the value of $\Delta$ - you don't need to memorize formula.



## Tests of $H_0 : \mu_1 = \mu_2$

Similar to CIs

**A.** Assume $\sigma_1 = \sigma_2$. Reject if $\dfrac{|\bar{Y}_1 - \bar{Y}_2|}{SE_p} \geq t_{\alpha/2;\, n_1+n_2-2}$.

**B.** Don't assume $\sigma_1 = \sigma_2$. Reject if $\dfrac{|\bar{Y}_1 - \bar{Y}_2|}{SE_{unp}} \geq t_{\alpha/2;\, \Delta}$.

One-sided tests are also available. See p. 46.

Now apply these procedures to our advertising data:

## **t-tests** (JMP printout)

**A.** Assume $\sigma_1 = \sigma_2$

| | Difference | t-Test | t-Test DF | Prob>\|t\| |
|---|---|---|---|---|
| Estimate | 2.49 | **0.343** | 26 | **0.7341** |
| Std Error | 7.26 | | | |
| Lower 95% | -12.43 | | | |
| Upper 95% | 17.42 | | | |

Assuming equal variances

**B.** Don't assume $\sigma_1 = \sigma_2$

Welch Anova testing Means Equal, allowing Std's Not Equal

| F Ratio | DF Num | DF Den* | Prob>F* |
|---|---|---|---|
| 0.1164 | 1 | 24.66 | **0.7359** |

**t-Test**

**0.341**

*Here, "Prob>F" is the P-value. Also, $\Delta$ = "DFDen" = 24.66. Note that in such a table $.341^2 = t^2 = F = .1164$

In JMP: Use Analyze → Fit Y by X platform. For test A use "Means/ANOVA/pooled t" button. For test B use "UnEqual Variances" button and ignore output you don't need.



## **Confidence Intervals** (JMP printout)

**A.** Assume $\sigma_1 = \sigma_2$

Std Error = 7.26 (from **t-Test** table)

DF = $n_1 + n_2 - 2 = 13 + 15 - 2 = 26$ (or see DF in **t-Test** table)

For 95% CI $t_{\alpha/2;26} = 2.056$ (from Table B.2, *or from JMP*)

95% CI: $2.49 \pm 2.056 \times 7.26 = 2.49 \pm 14.93$

**OR** read the result from JMP: (-12.43, 17.42)

You can get other %-age CIs by using Table B.2 or in JMP with the button, "Set Alpha Level".

**B.** Don't assume $\sigma_1 = \sigma_2$

Std Error: This isn't directly available in JMP. You need to calculate from the formula and Summary Table or work backward from $\bar{Y}_1 - \bar{Y}_2$ in Summary Table via

$$SE = \frac{|\bar{Y}_1 - \bar{Y}_2|}{Welch's\ \mathbf{t}} = \frac{2.49}{.341} = 7.31.$$

DF: $\Delta = 24.66$ [$\geq$ **24**]

$t_{\alpha/2;\mathbf{24}} = 2.064$ (from Table B.2). Or $t_{\alpha/2;\mathbf{24.66}} = 2.061$ from JMP

95% CI: $2.49 \pm 2.06 \times 7.31 = 2.49 \pm 14.96$

__If__ exact t-values are not available, use $t_{\alpha/2;DF} \approx 2$ so long as DF $\geq$ 20.

## Matched Pairs analysis

**Actually**, the sampling method here for choosing stores was more sophisticated than the random sampling scheme assumed above.

What was actually done is that **14 *pairs*** of stores were chosen - 28 stores in all. The two stores in each pair were pretty well matched as to overall size and usual weekly sales and demographic patterns of customers.[1]

One store in each pair was designated to receive the ad campaign; the other to not receive it.

Unfortunately, for one pair the designated ad-store did not receive its advertising materials in time to run the campaign. Thus 13 stores actually ran the campaign and 15 did not.

1. This type of Matched-Pairs design, where the matching is performed by the statistician/analyst on the basis of additional variables (such as size and usual weekly sales) is called a **Case-Control** study.



Here is the complete data:

| Pair | With | Without | Difference |
|---|---|---|---|
| 1 | 67.2 | 65.3 | 1.9 |
| 2 | 59.4 | 54.7 | 4.7 |
| 3 | 80.1 | 81.3 | -1.2 |
| 4 | 47.6 | 39.8 | 7.8 |
| 5 | 97.8 | 92.5 | 5.3 |
| 6 | 38.4 | 37.9 | 0.5 |
| 7 | 57.3 | 52.4 | 4.9 |
| 8 | 75.2 | 69.9 | 5.3 |
| 9 | 94.7 | 89 | 5.7 |
| 10 | 64.3 | 58.4 | 5.9 |
| 11 | 31.7 | 33 | -1.3 |
| 12 | 49.3 | 41.7 | 7.6 |
| 13 | 54 | 53.6 | 0.4 |
| 14 | * | 63.2 | * |
| 14 | * | 72.6 | * |

## Matched Pairs Test

The matched-pair method tests whether $\mu_1 = \mu_2$ ***after taking into account store size**, **etc.***

Look at the pairwise differences, and use them to test

$$H_0: \mu_{difference} = 0 \text{ versus } H_a: \mu_{difference} \neq 0.$$

In doing this test we'll have to ignore the results from the two stores in Pair 14, since neither of them had the ad campaign.

Given the sample mean and standard deviation of the differences *you* should be able to compute the relevant **t-Test** statistic, the p-value, and the $100(1-\alpha)\%$ CI.

But JMP will (also) do this for you:



## JMP Output

Variable: Differences

**Moments**

| | |
|---|---|
| Mean | 3.66 |
| Std Dev | 3.19 |
| Std Error Mean | 0.88 |
| Upper 95% Mean | 5.58 |
| Lower 95% Mean | 1.73 |
| N | 13 |

**Test Mean=value**

| | |
|---|---|
| Hypothesized Value | 0 |
| Actual Estimate | 3.65 |

**t Test**

| | |
|---|---|
| Test Statistic | 4.14 |
| Prob > \|t\| | 0.0014 |

**Conclusion**: **Taking into account the store sizes, etc**. an ad campaign **does** improve store sales (with P-value .0014) by between $1,730 and $5,580 per week (95% CI).

## Notes:

- Which test is preferable?

**A.** For test that assumes $\sigma_1 = \sigma_2$:

If assumption is correct test is "**Exact**" – *ie*, $P_{H_0}(\text{Rej } H_0) = \alpha$, exactly.

If assumption is correct it has (slightly) higher power than alternate test B.
This is the historically more familiar test.
All our regression procedures and ANOVA procedures make this type of assumption.

**B.** For test that does not assume $\sigma_1 = \sigma_2$:

Test is only approximately level α - *ie*, $P_{H_0}(\text{Rej } H_0) \approx \alpha$; But the approximation is good unless $\min(n_1, n_2)$ is not large and $\min(\sigma_1/\sigma_2, \sigma_2/\sigma_1)$ is not near 1.

Even when $\sigma_1 = \sigma_2$, the power is almost as good as test **A**.

No assumption is needed about $\sigma_1, \sigma_2$.

HENCE: With special exceptions test B is usually preferable in practice,
But test A is more familiar and can safely be used whenever the assumption $\sigma_1 = \sigma_2$ seems reasonably close to being correct. $\max(s_1^2/s_2^2, s_2^2/s_1^2) > 3$ is a **Danger!** sign.

**C.** If there is even a mildly sensible matching, the Matched-Pairs test is preferable.
It does not require the assumption that $\sigma_1 = \sigma_2$. It will be exact (under the normal distribution assumption). With any moderately sensible matching it will have higher power than tests A and B.



- How big is the DF number Δ?

$$\min(n_1, n_2) - 1 \le \Delta \le n_1 + n_2 - 2$$

with the bounding values occurring as $s_1/s_2 \to 0$ or $\infty$ and $s_1/s_2 = 1$, resp.

SO,

$\min(n_1, n_2) - 1$ is always a conservatively suitable choice for Δ.

- How far from α is the actual level of the test, B, that does not assume $\sigma_1 = \sigma_2$?

General Answer: Not far at all

Specific answer depends on $n_1, n_2\, n_1, n_2$ and on $\sigma_2^2/\sigma_1^2$. Here is a plot showing approximate values of the true level of the test for $n_1 = 10, n_2 = 20$ and for various values of $\log\left[\sigma_2^2/\sigma_1^2\right]$. {Labeled on the plot, for convenience, as "$\log \sigma_2^2$".}

a. The points on the plot are Monte-Carlo 'estimates' of the true value. They are close to the true value, but contain some random error.

b. The smooth-ish curve on the plot provides a better approximation to the true value of α = "Prob of rejection under $H_0$".

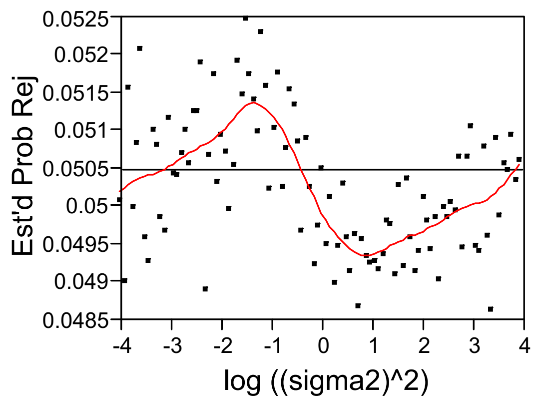
c. The x-axis on the plot is $\log\left[\sigma_2^2/\sigma_1^2\right]$. It runs from -4 to +4. This means that $\sigma_2^2/\sigma_1^2$ runs from 0.018 to 50.4, which is a quite wide range of values.

d. The estimated values of α are all within the range (.0485, .0525); hence are all pretty close to the nominal value of 0.05.

Plot of the Monte-Carlo estimates of the true $\alpha$

for a test at desired level $\alpha = 0.05$, not assuming $\sigma_1 = \sigma_2$.

Here $n_1 = 10, n_2 = 20$, $\sigma_1^2 = 1$.




Here's one more plot; this time for smaller sample sizes



Plot of the Monte-Carlo estimates of the true $\alpha$

for a test at desired level $\alpha = 0.05$, not assuming $\sigma_1 = \sigma_2$.

Here $n_1 = 4, n_2 = 9$, $\sigma_1^2 = 1$.