Comparison of Two Means

Comparing Two Means

Two Normally-Distributed Populations:

- $X^{(1)} \sim N(\mu_1, \sigma_1^2)$,
- $X^{(2)} \sim N(\mu_2, \sigma_2^2)$.

Goal: compare μ_1 and μ_2.

Three Basic Cases:

- σ_1^2 and σ_2^2 are known
- σ_1^2 and σ_2^2 are unknown but $\sigma_1^2 = \sigma_2^2$
- σ_1^2 and σ_2^2 are unknown and not necessarily equal

Also:

- paired comparisons
- non-parametric tests

A Point Estimator for the Difference of Means

We take random samples $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$ of sizes n_1 and n_2 from the populations, we can find a point estimator for the difference of the two means

$$\widehat{\mu_1 - \mu_2} := \widehat{\mu}_1 - \widehat{\mu}_2 = \overline{X}^{(1)} - \overline{X}^{(2)}.$$

Since

$$\overline{X}^{(1)} \sim N(\mu_1, \sigma_1^2/n_1), \qquad \overline{X}^{(2)} \sim N(\mu_2, \sigma_2^2/n_2),$$

we see that $\overline{X}_1 - \overline{X}_2$ is normal with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$, i.e.,

$$\frac{\overline{X}^{(1)} - \overline{X}^{(2)} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

is a standard normal random variable.

Neyman-Pearson Test with Variances Known

We may use this result to obtain confidence intervals for the difference of means and to conduct hypothesis tests.

21.1. Example. The plant manager at an orange juice canning facility is interested in comparing the performance of two different production lines in her plant. As line number 1 is relatively new, she suspects that its output in number of cases per day is greater than the number of cases produced by the older line 2.

She sets up the hypotheses

$$H_0\colon \mu_1 \le \mu_2, \qquad\qquad H_1\colon \mu_1 > \mu_2 + 10\,\text{cases}\,.$$

She decides to use $\alpha = 5\%$. The test statistic is

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

and H_0 will be rejected if this number is greater than $z_{0.05} = 1.645$.

Neyman-Pearson Test with Variances Known

From experience with operating this type of equipment it is known that $\sigma_1^2 = 40$ and $\sigma_2^2 = 50$.

Ten days of data are selected at random for each line, for which it is found that $\overline{x}^{(1)} = 824.9$ cases per day and $\overline{x}^{(2)} = 818.6$ cases per day.

The value of the test statistic is then calculated to be

$$Z = \frac{824.9 - 818.6}{\sqrt{40/10 + 50/10}} = 2.10.$$

Since $Z > 1.645$ we reject H_0 at a 5% level of significance. The alternative hypothesis H_1 is accepted.

The plant manager concludes that the new production line produces 10 cases per day more than the older line (and may decide to replace more of the older lines as a consequence).

OC Curves for Variances Known

We can also use the OC curves for the normal distribution to find power and sample size for a test. In that case, we use

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

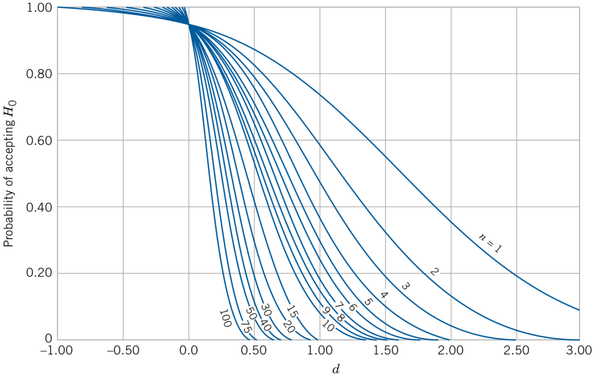
with $n = n_1 = n_2$ (equal sample sizes).

If $n_1 \neq n_2$, the table is used with the ***equivalent sample size***

$$n = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}.$$

OC Curves for Variances Known

21.2. Example. Continuing from Example 21.1, if H_1 is true, we want to find the sample sizes (number of days) required to detect this difference with a probability of 0.90.



We have $d = 10/\sqrt{40+50} = 1.05$ and using the chart for $\alpha = 0.05$ (one-sided) we find $n = n_1 = n_2 = 9$.

Confidence Interval for the Difference of Means

21.3. Example. Using the data of Example 21.1, a 95% confidence interval for the difference in mean production is

$$\begin{aligned}\mu_1 - \mu_2 &= \overline{x}^{(1)} - \overline{x}^{(2)} \pm z_\alpha \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \\ &= 824.9 - 818.6 \pm 1.645\sqrt{40/10 + 50/10} \\ &= 6.3 \pm 4.9\end{aligned}$$

Note that zero is not in this confidence interval, which is expected since $H_0\colon \mu_1 \le \mu_2$ was rejected.

Comparing Two Means - Equal Variances

Now suppose that the variances are equal but unknown,

$$\sigma_1^2 = \sigma_2^2 =: \sigma^2.$$

Then

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n_1 + 1/n_2)}}.$$

is standard normal

Similarly to (19.1), we define the ***pooled estimator for the variance***

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}. \tag{21.1}$$

Comparing Two Means - Equal Variances

It is immediately clear that

$$X^2_{n_1+n_2-2} = \frac{(n_1+n_2-2)S_p^2}{\sigma^2} = \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2}$$

follows a chi-squared distribution with n_1+n_2-2 degrees of freedom.

Furthermore,

$$\begin{aligned} T_{n_1+n_2-2} &= \frac{Z}{\sqrt{X^2_{n_1+n_2-2}/(n_1+n_2-2)}} \\ &= \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} \end{aligned}$$

follows a T-distribution with n_1+n_2-2 degrees of freedom.

Confidence Interval for the Difference of Means

We immediately obtain the following $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$,

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2, n_1+n_2-2}\sqrt{S_p^2(1/n_1 + 1/n_2)},$$

where $t_{\alpha/2, n_1+n_2-2}$ is defined in (13.5).

21.4. Example. In a batch chemical process used for etching circuit boards, two different catalysts are being compared to determine whether they require different emersion times for removal of identical quantities of photo-resistant material.

Twelve batches were run with catalyst 1, resulting in a sample mean emersion time of $\overline{x}_1 = 24.6$ minutes and a sample standard deviation of $s_1 = 0.85$ minutes. Fifteen batches were run with catalyst 2, resulting in a mean emersion time of $\overline{x}_2 = 22.1$ minutes and a standard deviation of $s_2 = 0.98$ minutes.

Confidence Interval for the Difference of Means

We will find a 95% confidence interval on the difference in means $\mu_1 - \mu_2$ assuming that the variances of the two populations are equal. The pooled estimate for the variance gives

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 0.8557$$

so $s_p = 0.925$. Since $t_{0.025,25} = 2.060$, we obtain

$$\mu_1 - \mu_2 = (2.5 \pm 0.74)\,\text{minutes}$$

Student's T-Test for Equal Variances

21.5. Student's T-Test for Equal Variances. Suppose two random samples of sizes n_1 and n_2 from two normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ are given.

Denote by $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$ the means of the two samples and let S_p^2 be the pooled sample variance (21.1). Let $(\mu_1 - \mu_2)_0$ be a null value for the difference $\mu_1 - \mu_2$. Then the test based on the statistic

$$T_{n_1+n_2-2} = \frac{(\overline{X}^{(1)} - \overline{X}^{(2)}) - (\mu_1 - \mu_2)_0}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$$

is called a ***Student's (pooled) test for equality of means***.

We reject at significance level α

- $H_0\colon \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$ if $|T_{n_1+n_2-2}| > t_{\alpha/2, n_1+n_2-2}$,
- $H_0\colon \mu_1 - \mu_2 \le (\mu_1 - \mu_2)_0$ if $T_{n_1+n_2-2} > t_{\alpha, n_1+n_2-2}$,
- $H_0\colon \mu_1 - \mu_2 \ge (\mu_1 - \mu_2)_0$ if $T_{n_1+n_2-2} < -t_{\alpha, n_1+n_2-2}$.

Student's T-Test for Equal Variances

21.6. Example. Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, catalyst 1 is currently in use, but catalyst 2 is acceptable. Since catalyst 2 is cheaper, if it does not change the process yield significantly, it should be adopted.

We decide to test the hypotheses

$$H_0\colon \mu_2 \geq \mu_1, \qquad\qquad H_1\colon \mu_2 \leq \mu_1 - 3\%.$$

From experience with this type of chemical process, the yield follows a normal distribution and the variance of the yield is independent of the catalyst used.

We therefore conduct a Student T-test with $\alpha = 5\%$ and choose sample sizes $n_1 = n_2 = 8$. Then the critical value of the test statistic is $t_{0.05,14} = 1.761$.

Student's T-Test for Equal Variances

Pilot data yields

$$\overline{x}_1 = 93.75\%, \qquad s_1^2 = 3.89\%^2, \qquad \overline{x}_2 = 91.73\%, \qquad s_2^2 = 4.02\%^2.$$

Then $s_p^2 = 3.96\%^2$ and the test statistic is

$$\frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = 2.03.$$

Since this is greater than the critical value, we reject H_0 and accept H_1.

We conclude that catalyst 2 induces a significantly lower yield (by 3%) than catalyst 1.

OC Curves for Equal Variances

In the case of equal variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and equal sample sizes $n_1 = n_2 = n$, we can use the usual OC curves for the T-test with

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma}.$$

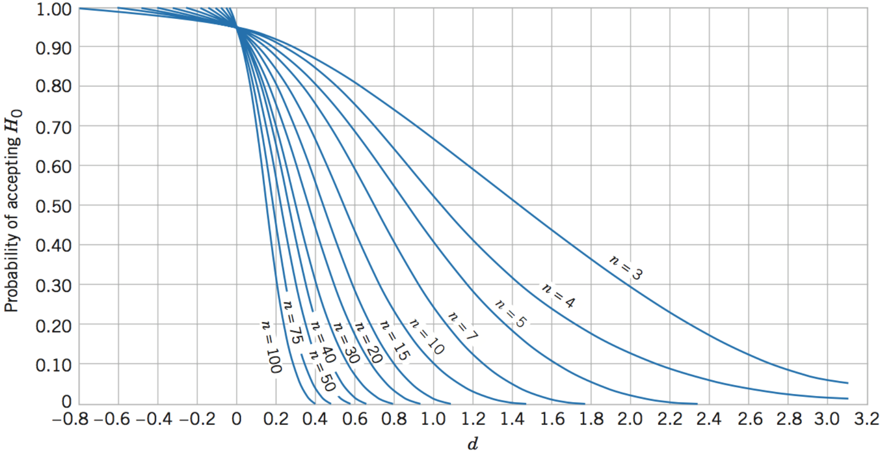
However, we must use the ***modified sample size*** $n^* = 2n - 1$ when reading the charts. As before, when σ is unknown, we must either use an estimate or express the deviation in terms of σ.

21.7. Example. In setting up the experiment of Example 21.6 it was desired that the power should be at least 0.85. What sample size was required?

We use $\alpha = 0.05$ and the previously determined $s_p = 1.99$ as an estimate for the common standard deviation σ.

OC Curves for Equal Variances

Then $d = \delta/(2\sigma) = 3/(2 \cdot 1.99) = 0.75$ and $\beta = 1 - 0.85 = 0.15$.



The chart gives $n^* = 15$, so $n = (n^* + 1)/2 = 8$ was sufficient.

A Warning Regarding Pre-Testing

The previous discussion of Student's T-test for comparison of means made two assumptions:

- Both random variables follow normal distributions.
- Both random variables have equal variances σ^2.

Comparing the means of two populations is a very common procedure in many applied sciences. In such cases, there is a temptation to

(i) Collect data.
(ii) Perform pre-tests on the data (e.g., test for equality of variances or test for normality)
(iii) Then perform the comparison of means test depending on the result of the pre-test.

This is not recommended!

A Warning Regarding Pre-Testing

Performing such pre-tests and then conditionally on the results using some other test ***on the same data*** will ***invalidate the P-value*** of the comparison of means test.

It is fine to test for normality, equality of variances or other properties and then to ***gather new data for a comparison of means test.*** But using the same data creates serious problems.

Literature:

- Rasch, D., Kubinger, K. and Moder, K. ***The two-sample T test: Pre-testing its assumptions does not pay off.*** Stat. Pap. 52 (2011).
- Rochon, J., Gondan, M. and Kieser, M. ***To test or not to test: Preliminary assessment of normality when comparing two independent samples.*** BMC Med Res Methodol 12, 81 (2012).
- Zimmerman, D. W. ***A note on preliminary tests of equality of variances.*** Br J Math Stat Psychol. 57 (2004).

Populations with Unequal Variances

We now consider the case of two normal populations with unequal variances. Recall that

$$\frac{\overline{X}^{(1)} - \overline{X}^{(2)} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

follows a standard normal distribution.

Now if the variances of the populations are not equal and unknown to us, we are faced with estimating the variance:

$$\left(\widehat{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}.$$

The main problem is that the distribution of the right-hand side is unknown.

The Welch-Satterthwaite Approximation

21.8. Welch-Satterthwaite Relation. Let $X^{(1)}, \dots, X^{(k)}$ be k independent normally distributed random variables with variances $\sigma_1^2, \dots, \sigma_k^2$.

Let $s_1^2, \dots, s_k^2$ be sample variances based on samples of sizes $n_1, \dots, n_k$ from the k populations, respectively. Let $\lambda_1, \dots, \lambda_k > 0$ be positive real numbers and define

$$\gamma := \frac{(\lambda_1 s_1^2 + \cdots + \lambda_k s_k^2)^2}{\displaystyle\sum_{i=1}^{k} \frac{(\lambda_i s_i^2)^2}{n_i - 1}}.$$

Then

$$\gamma \cdot \frac{\lambda_1 s_1^2 + \lambda_2 s_2^2 + \cdots + \lambda_k s_k^2}{\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2 + \cdots + \lambda_k \sigma_k^2}$$

follows ***approximately*** a chi-squared distribution with γ degrees of freedom.

The Welch-Satterthwaite Approximation

We are interested in the case $k = 2$, $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$. Then

$$\gamma = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}. \tag{21.2}$$

and

$$\gamma \cdot \frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

follows approximately a chi-squared distribution with γ degrees of freedom. It is then easy to see that

$$T_\gamma = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

follows a T-distribution with γ degrees of freedom.

Welch's T-Test for Unequal Variances

21.9. Welch's T-Test for Unequal Variances. Suppose two random samples of sizes n_1 and n_2 from two normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ are given.

Denote by $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$ the means of the two samples and let γ given by (21.2). Let $(\mu_1 - \mu_2)_0$ be a null value for the difference $\mu_1 - \mu_2$. Then the test based on the statistic

$$T_\gamma = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

is called a ***Welch's (pooled) test for equality of means***. We reject at significance level α

- $H_0\colon \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$ if $|T_\gamma| > t_{\alpha/2,\gamma}$,
- $H_0\colon \mu_1 - \mu_2 \le (\mu_1 - \mu_2)_0$ if $T_\gamma > t_{\alpha,\gamma}$,
- $H_0\colon \mu_1 - \mu_2 \ge (\mu_1 - \mu_2)_0$ if $T_\gamma < -t_{\alpha,\gamma}$.

Welch's T-Test for Unequal Variances

21.10. Remarks.

- In practice, we ***round γ down*** to the nearest integer.
- One disadvantage of unequal variances is that power calculations are much more difficult. There are no simple OC curves for Welch's test.
- As remarked earlier, it is not a good idea to pre-test for equal variances and then make a decision whether to use Student's or Welch's test. In fact, current recommendations are to ***always use Welch's test***. (This is different from what you find in most textbooks. See, for example, the literature below.)

  The reason is that Welch's test is only slightly less powerful that Student's test even if the variances are equal. if they are unequal, Student's test is very unreliable.

Literature: The blog article at `http://daniellakens.blogspot.com/2015/01/always-use-welchs-t-test-instead-of.html` and the author's paper cited there.