STAT 509: Statistics for Engineers

Chapter 10: Statistical Inference for Two Samples

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

Fall 2020

Learning Objectives:

1. Structure comparative experiments involving two samples as hypothesis tests
2. Test hypotheses and construct confidence intervals on the difference in means of two normal distributions
3. Test hypotheses and construct confidence intervals on the ratio of the variances or standard deviations of two normal distributions
4. Test hypotheses and construct confidence intervals on the difference in two population proportions
5. Use the P-value approach for making decisions in hypotheses tests
6. Explain and use the relationship between confidence intervals and hypothesis tests

Inference on the Difference in Means of Two Normal Distributions, Variances Known

The previous two chapters presented hypothesis tests and confidence intervals for a single population parameter (the mean μ, the variance σ^2, or a proportion p). This chapter extends those results to the case of two independent populations.

Most of the practical applications of the procedures in this chapter arise in the context of simple **comparative experiments** in which the objective is to study the difference in the parameters of the two populations.

Inference on the Difference in Means of Two Normal Distributions, Variances Known

Engineers and scientists are often interested in comparing two different conditions to determine whether either condition produces a significant effect on the response that is observed. These conditions are sometimes called **treatments**.

For example, a product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. A study was conducted to determine whether the new formulation results in a significant effect—reducing drying time. In this situation, the product developer (the experimenter) randomly assigned 10 test specimens to one formulation and 10 test specimens to the other formulation. Then the paints were applied to the test specimens in random order until all 20 specimens were painted. This is an example of a **completely randomized experiment**.

Inference on the Difference in Means of Two Normal Distributions, Variances Known

When statistical significance is observed in a randomized experiment, the experimenter can be confident in the conclusion that the difference in treatments resulted in the difference in response. That is, we can be confident that a **cause-and-effect** relationship has been found.

Sometimes the objects to be used in the comparison are not assigned at random to the treatments. A study, done in Finland, tracked 1931 men for 5 years and showed a statistically significant effect of increasing iron levels on the incidence of heart attacks. In this study, the comparison was not performed by randomly selecting a sample of men and then assigning some to a "low iron level" treatment and the others to a "high iron level" treatment. The researchers just tracked the subjects over time. This type of study is called an **observational study**.

It is difficult to identify causality in observational studies because the observed statistically significant difference in response for the two groups may be due to some other underlying factor (or group of factors) that was not equalized by randomization and not due to the treatments. For example, the difference in heart attack risk could be attributable to the difference in iron levels or to other underlying factors that form a reasonable explanation for the observed results—such as cholesterol levels or hypertension.

In this chapter, we assume that we have two populations. Based on a random sample from each, we conduct two-sample inference:

1. $X_{11}, \ldots, X_{1n_1}$ is a random sample of size n_1 from population 1.
2. $X_{21}, \ldots, X_{2n_2}$ is a random sample of size n_2 from population 2.

The n_1 and n_2 could be different.

Inference on the Difference in Means of Two Normal Distributions, Variances Known

We start with the **difference in means** $\mu_1 - \mu_2$ of two **normal** distributions where the variances σ_1^2 and σ_2^2 are **known**. We consider two types of inferences:

(1) Constructing a $100(1-\alpha)\%$ confidence interval

- twosample.Zinterval(level=?, sigma1=?,sigma2=?,sample1=?,sample2=?)
- twosample.Zinterval(level=?, sigma1=?,sigma2=?,barx1=?,barx2=?,n1=?,n2=?)

(2) and testing the null hypothesis $H_0 : \mu_1 - \mu_0 = \Delta_0$ for a hypothesized value Δ_0 against a two-tailed ($\neq$), left-tailed ($<$), or right-tailed ($>$) alternative hypothesis.

- twosample.Ztest(Delta0=?,H1=?,alpha=?,sigma1=?, sigma2=?,sample1=?,sample2=?)
- twosample.Ztest(Delta0=?,H1=?,alpha=?,sigma1=?, sigma2=?,barx1=?,barx2=?,n1=?,n2=?)

Inference on the Difference in Means of Two Normal Distributions, Variances Known

Same as the one-sample case, the types of inferences both start with a point estimator of $\mu_1 - \mu_2$, which is

$$\bar{X}_{1n_1} - \bar{X}_{2n_2} = n_1^{-1} \sum_{i=1}^{n_1} X_{1i} - n_2^{-1} \sum_{i=1}^{n_2} X_{2i}.$$

And we know that

$$Z = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

Thus, the two-tailed $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ comes from

$$-z_{\alpha/2} \leq Z = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{\alpha/2}.$$

Inference on the Difference in Means of Two Normal Distributions, Variances Known

When testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ for a hypothesized value μ_0,
The test statistic is $Z_0 = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.
Based on samples $x_{11}, \ldots, x_{1n_1}$ and $x_{21}, \ldots, x_{2n_2}$, the observed test statistic is $z_0 = \frac{\bar{x}_{1n_1} - \bar{x}_{2n_2} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.
If the alternative hypothesis is $H_1 : \mu_1 - \mu_2 \neq \Delta_0$. At significance level α, we reject H_0 if one of the following holds:

- $|z_0| > z_{\alpha/2}$ (critical-value approach),
- $\mu_0 \notin \left[\bar{x}_{1n_1} - \bar{x}_{2n_2} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$ (confidence-interval approach),
- the P-value $= 2[1 - P(Z \leq |z_0|)] < \alpha$, where $Z \sim N(0,1)$ (P-value approach).

Tests on the Mean of a Normal Distribution, Variance Known

If the alternative hypothesis is $H_1 : \mu_1 - \mu_2 > \Delta_0$. At significance level α, we reject H_0 if one of the following holds:

- $z_0 > z_\alpha$ (critical-value approach),
- $\mu_0 \notin \left[\bar{x}_{1n_1} - \bar{x}_{2n_2} - z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \infty\right)$ (confidence-interval approach),
- the P-value $= P(Z > z_0) < \alpha$, where $Z \sim N(0,1)$ (P-value approach).

If the alternative hypothesis is $H_1 : \mu_1 - \mu_2 < \Delta_0$. At significance level α, we reject H_0 if one of the following holds:

- $z_0 < -z_\alpha$ (critical-value approach),
- $\mu_0 \notin \left(-\infty, \bar{x}_{1n_1} - \bar{x}_{2n_2} + z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$ (confidence-interval approach),
- the P-value $= P(Z \leq z_0) < \alpha$, where $Z \sim N(0,1)$ (P-value approach).

Example 1

Tensile strength tests were performed on two different grades of aluminum spars used in manufacturing the wing of a commercial transport aircraft. From past experience with the spar manufacturing process and the testing procedure, the standard deviations of tensile strengths are assumed to be known. The data obtained are as follows: $n_1 = 10$, $\bar{x}_1 = 87.6$, $\sigma_1 = 1$, $n_2 = 12$, $\bar{x}_2 = 74.5$, and $\sigma_2 = 1.5$. If μ_1 and μ_2 denote the true mean tensile strengths for the two grades of spars, find a 90% confidence interval on the difference in mean strength. Assume normality.

solution: Normality and known variances.

```
twosample.Zinterval(level=0.9, sigma1=1,sigma2=1.5,
                    barx1=87.6,barx2=74.5,n1=10,n2=12)
A 90% two-sided confidence interval for
the difference in population means is [ 12.21805 , 13.98195 ]
A 90% upper-confidence bound for the population mean is 13.78716
A 90% lower-confidence bound for the population mean is 12.41284
```

Conclusion: Based on the data, a 90% confidence interval on $\mu_1 - \mu_2$ is $[12.2181, 13.982]$.

Example 2

A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

Example 2

solution: Normality and known variances with $\sigma_1 = \sigma_2 = 8$. We are testing $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 > \mu_2$. Then $\Delta_0 = 0$ and we have a right-tailed alternative.

```
twosample.Ztest(Delta0=0,H1="right",alpha=0.05, sigma1=8,sigma2=8,
                barx1=121,barx2=112,n1=10,n2=10)
H1 is right-tailed. The results are:

1. Test statistic z0 is 2.515576 , z_alpha is 1.644854 .
Because z0>z_alpha, we reject H0 at significance level 0.05

2. A 95% one-sided confidence interval for the population mean is
[ 3.115193 , Inf ) which does not contain the hypothesized value Delta0= 0 ,
so we reject H0 at significance level 0.05

3. The P-value is 0.005941895 which is smaller than alpha= 0.05 ,
so we reject H0 at significance level 0.05
```

Conclusion: at significance level 0.05, the data provide sufficient evidence to reject H_0.

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=yes or no

We then consider the **difference in means** $\mu_1 - \mu_2$ of two **normal** distributions where the variances σ_1^2 and σ_2^2 are **unknown**. In order to have a better power, we separate the unknown variances into two cases

Case I: pooled=yes: We do not know σ_1^2 or σ_2^2, but there is information for us to conclude $\sigma_1^2 = \sigma_2^2$.

Case I: pooled=no: We do not know σ_1^2 or σ_2^2, and there is no information for us to conclude $\sigma_1^2 = \sigma_2^2$.

If we are able to conclude $\sigma_1^2 = \sigma_2^2$ (Case I), this extra piece of information can help us gain a better power for testing whether $H_0 : \mu_1 - \mu_2 = \Delta_0$ holds.

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=yes

We then consider the **difference in means** $\mu_1 - \mu_2$ of two **normal** distributions where the variances σ_1^2 and σ_2^2 are **unknown** but we can conclude $\sigma_1^2 = \sigma_2^2$ (Pooled=yes). We consider two types of inferences:

(1) Constructing a $100(1-\alpha)\%$ confidence interval
- twosample.Tinterval(level=?, pooled=yes, sample1=?,sample2=?)
- twosample.Tinterval(level=?, pooled=yes, barx1=?,barx2=?,n1=?,n2=?,s1=?,s2=?)

(2) and testing the null hypothesis $H_0 : \mu_1 - \mu_0 = \Delta_0$ for a hypothesized value Δ_0 against a two-tailed ($\neq$), left-tailed ($<$), or right-tailed ($>$) alternative hypothesis.
- twosample.Ttest(Delta0=?,H1=?,alpha=?,pooled=yes, sample1=?,sample2=?)
- twosample.Ttest(Delta0=?,H1=?,alpha=?,pooled=yes, barx1=?,barx2=?,n1=?,n2=?,s1=?,s2=?)

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=yes

The point estimator of $\mu_1 - \mu_2$ is still $\bar{X}_{1n_1} - \bar{X}_{2n_2}$. If $\sigma_1^2 = \sigma_2^2$ holds, we let the common variance be σ^2. Then we can combine the two samples together to get a pooled estimator of σ^2, denoted by S_p^2,

$$S_p^2 = \frac{(n_1 - 1)S_{1n_1}^2 + (n_2 - 1)S_{2n_2}^2}{n_1 + n_2 - 2}.$$

Then

$$T = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2).$$

Thus, the two-tailed $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ comes from

$$-t_{n_1+n_2-2,\alpha/2} \leq T = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{n_1+n_2-2,\alpha/2}.$$

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=yes

When testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ for a hypothesized value μ_0,
The test statistic is $T_0 = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - \Delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$.

Based on samples $x_{11}, \dots, x_{1n_1}$ and $x_{21}, \dots, x_{2n_2}$, the observed test statistic is $t_0 = \frac{\bar{x}_{1n_1} - \bar{x}_{2n_2} - \Delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$.

- If $H_1 : \mu_1 - \mu_2 \neq \Delta_0$. At significance level α, we reject H_0 if $|t_0| > t_{n_1+n_2-2,\alpha/2}$ (critical-value approach),
- If $H_1 : \mu_1 - \mu_2 < \Delta_0$. At significance level α, we reject H_0 if $t_0 < -t_{n_1+n_2-2,\alpha}$ (critical-value approach),
- If $H_1 : \mu_1 - \mu_2 > \Delta_0$. At significance level α, we reject H_0 if $t_0 > t_{n_1+n_2-2,\alpha}$ (critical-value approach).

We also have the confidence-interval approach and the P-value approach.

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=no

We then consider the **difference in means** $\mu_1 - \mu_2$ of two **normal** distributions where the variances σ_1^2 and σ_2^2 are **unknown** but we cannot conclude $\sigma_1^2 = \sigma_2^2$ (Pooled=no). We consider two types of inferences:

(1) Constructing a $100(1-\alpha)\%$ confidence interval
- twosample.Tinterval(level=?, pooled=no, sample1=?,sample2=?)
- twosample.Tinterval(level=?, pooled=no, barx1=?,barx2=?,n1=?,n2=?,s1=?,s2=?)

(2) and testing the null hypothesis $H_0 : \mu_1 - \mu_0 = \Delta_0$ for a hypothesized value Δ_0 against a two-tailed ($\neq$), left-tailed ($<$), or right-tailed ($>$) alternative hypothesis.
- twosample.Ttest(Delta0=?,H1=?,alpha=?,pooled=no, sample1=?,sample2=?)
- twosample.Ttest(Delta0=?,H1=?,alpha=?,pooled=no, barx1=?,barx2=?,n1=?,n2=?,s1=?,s2=?)

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=no

The point estimator of $\mu_1 - \mu_2$ is still $\bar{X}_{1n_1} - \bar{X}_{2n_2}$. If $\sigma_1^2 = \sigma_2^2$ does not hold, we estimate them separately by the sample variances S_1^2 and S_2^2. Then

$$T = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(v),$$

where

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

Thus, the two-tailed $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ comes from

$$-t_{v,\alpha/2} \leq T = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{v,\alpha/2}.$$

Inference on the Difference in Means of Two Normal Distributions, Variances Unknown, pooled=no

When testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ for a hypothesized value μ_0,
The test statistic is $T_0 = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$.

Based on samples $x_{11}, \ldots, x_{1n_1}$ and $x_{21}, \ldots, x_{2n_2}$, the observed test statistic is $t_0 = \frac{\bar{x}_{1n_1} - \bar{x}_{2n_2} - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.

- If $H_1 : \mu_1 - \mu_2 \neq \Delta_0$. At significance level α, we reject H_0 if $|t_0| > t_{v,\alpha/2}$ (critical-value approach),
- If $H_1 : \mu_1 - \mu_2 < \Delta_0$. At significance level α, we reject H_0 if $t_0 < -t_{v,\alpha}$ (critical-value approach),
- If $H_1 : \mu_1 - \mu_2 > \Delta_0$. At significance level α, we reject H_0 if $t_0 > t_{v,\alpha}$ (critical-value approach).

We also have the confidence-interval approach and the P-value approach.

Example 3

Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, catalyst 1 is currently used; but catalyst 2 is acceptable. Because catalyst 2 is cheaper, it should be adopted, if it does not change the process yield. A test is run in the pilot plant and results in the data.

```
catalyst1=c(91.50,94.18,92.18,95.39,91.79,89.07,94.72,89.21)
catalyst2=c(89.19,90.95,90.46,93.21,97.19,97.04,91.07,92.75)
```

Use QQ-plot to check normality assumption. Then answer: Is there any difference in the mean yields? Use $\alpha = 0.05$, and assume equal variances.

Solution: We first check normality using QQ-plot, which can be done using

```
data.summary(catalyst1)
data.summary(catalyst2)
```

Example 3

Then conduct the test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ in which $\Delta_0 = 0$. We have normality checked, but we do not know the population variances. Thus we use two-sample T-tests. Furthermore, the question assumes equal variances, which requires the use of “pooled=yes"

```
twosample.Ttest(Delta0=0,H1="two",alpha=0.05, pooled="yes",
                sample1=catalyst1,sample2=catalyst2)
H1 is two-tailed. The results are:

1. Test statistic t0 is -0.3535909 , t_(v,alpha/2) is 2.144787 .
Because |t0|<=t_(v,alpha/2), we fail to reject H0 at significance level 0.05

2. A 95% two-tailed confidence interval for the population mean is
[ -3.373886 , 2.418886 ] which contains the hypothesized value Delta0= 0 ,
so we fail to reject H0 at significance level 0.05

3. The P-value is 0.7289136 which is not smaller than alpha= 0.05 ,
so we fail to reject H0 at significance level 0.05
```

Conclusion: at significance level 0.05, the data do not provide sufficient evidence to reject H_0.

Example 4

An article in Polymer Degradation and Stability (2006, Vol. 91) presented data from a nine-year aging study on S537 foam. Foam samples were compressed to 50% of their original thickness and stored at different temperatures for nine years. At the start of the experiment as well as during each year, sample thickness was measured, and the thicknesses of the eight samples at each storage condition were recorded. The data for two storage conditions follow.

```
C50=c(0.047, 0.060, 0.061, 0.064, 0.080, 0.090, 0.118, 0.165, 0.183)
C60=c(0.062, 0.105, 0.118, 0.137, 0.153, 0.197, 0.210, 0.250, 0.335)
```

(a) Is there evidence to claim that mean compression increases with the temperature at the storage condition? ($\alpha = 0.05$).

(b) Find a 95% confidence interval for the difference in the mean compression for the two temperatures.

(c) Is the value zero contained in the 95% confidence interval? Explain the connection with the conclusion reached in part (a).

(d) Do normal probability plots of compression indicate any violations of the assumptions for the tests and confidence interval that you performed?

Example 4

Solution: (a) Let mu_1 be mean compression at 50 Celsius and μ_2 be the one at 60 Celsius. We are testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$ ($\Delta = 0$ and left-tailed).

```
twosample.Ttest(Delta0=0,H1="left",alpha=0.05, pooled="no",
                sample1=C50,sample2=C60)
H1 is left-tailed. The results are:

1. Test statistic t0 is -2.409256 , t_(v,alpha) is 1.772046 .
Because t0<-t_(v,alpha), we reject H0 at significance level 0.05

2. A 95% one-sided confidence interval for the population mean is
( -Inf , -0.0205416 ] which does not contain the hypothesized value Delta0= 0 ,
so we reject H0 at significance level 0.05

3. The P-value is 0.01583664 which is smaller than alpha= 0.05 ,
so we reject H0 at significance level 0.05
```

Conclusion: at significance level 0.05, the data provide sufficient evidence to reject H_0.

Thus, there is evidence to claim that mean compression increases with the temperature at the storage condition.

Example 4

Solution: (b,c) see result 2 of part (a). Because we are focusing on a one-side comparison. We use the one-sided confidence interval $(-\infty, -0.0205]$ for $\mu_1 - \mu_2$. Because 0 is not contained it, we would reject H_0 in (a).
(d) Check normality:

```
data.summary(C50)
data.summary(C60)
```

We can see that normality are checked. We could not conclude whether this is pooled or not. It is important to learn how to test whether two population variances are equal (later).

Paired T-Test and T-interval

A special case of the two-sample T-tests occurs when the observations on the two populations of interest are collected in pairs.

For example, suppose that we are interested in comparing two different types of tips for a hardness-testing machine. This machine presses the tip into a metal specimen with a known force. By measuring the depth of the depression caused by the tip, the hardness of the specimen can be determined.
If several specimens were selected at random, half tested with tip 1, half tested with tip 2, and a two sample t-test (pooled or not) was applied, the results of the test could be erroneous.
The metal specimens could have been cut from bar stock that was produced in different heats, or they might not be homogeneous in some other way that might affect hardness. Then the observed difference in mean hardness readings for the two tip types also includes hardness differences in specimens.

Paired T-Test and T-interval

A more powerful experimental procedure is to collect the data in pairs—that is, to make two hardness readings on each specimen, one with each tip. The test procedure would then consist of analyzing the differences in hardness readings on each specimen. If there is no difference between tips, the mean of the differences should be zero. This test procedure is called the paired T-test.

Let $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ be a set of n paired observations for which we assume that the mean and variance of the population represented by X_1 are μ_1 and σ_1^2, the mean and variance of the population represented by X_2 are μ_2 and σ_2^2. Define the difference for each pair of observations as $D_j = X_{1j} - X_{2j}$, $j = 1, 2, \ldots, n$. Then D_j's are assumed to be **normally** distributed with mean

$$\mu_D = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2$$

and some variance σ_D^2 (remains **unknown**).

Paired T-Test and T-interval

So testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ is equivalent to testing

$$H_0 : \mu_D = \Delta_0,$$

which can be done using the one-sample T-test:

```
Ttest(mu0=Delta0, H1=?,alpha=?,sample=sample1-sample2)
Ttest(mu0=Delta0, H1=?,alpha=?,n=?,
        barx=mean(sample1-sample2),s=sd(sample1-sample2))
```

Confidence intervals for $\mu_D = \mu_1 - \mu_2$ from paired samples can be done using the one-sample T-interval:

```
Tinterval(level=?,sample=?)
Tinterval(level=?,n=?,
        barx=mean(sample1-sample2),s=sd(sample1-sample2))
```

Example 5

An article in the Journal of Strain Analysis [1983, Vol. 18(2)] reports a comparison of several methods for predicting the shear strength for steel plate girders. Data for two methods, the Karlsruhe and Lehigh procedures, when applied to nine specific girders, are shown as

```
   Girder=c(1,    2,    3,    4,    5,    6,    7,    8,    9    )
Karlsrube=c(1.186,1.151,1.322,1.229,1.200,1.402,1.365,1.537,1.559)
   Lehigh=c(1.061,0.992,1.063,1.062,1.065,1.178,1.037,1.086,1.052)
```

We wish to determine whether there is any difference (on the average) for the two methods.

Solution: Obviously, data were taken in pairs. We should use paired T-test to determine whether there is any difference on the average for the two methods. The hypotheses are $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$, which equal to $H_0 : \mu_D = 0$ vs $H_1 : \mu_D \neq 0$. The *alpha* is not told, we can simply choose 0.05 to compute the P-value.

```
data.summary(Karlsrube-Lehigh) #check normality
Ttest(mu0=0, H1="two",alpha=0.05,sample=Karlsrube-Lehigh)
3. The P-value is 0.000498523
```

Conclusion: at significance level $\alpha > 0.0005$, data provide sufficient evidence to reject H_0; i.e., there very likely exists difference (on the average) between the two methods.

Inference on the Variances of Two Normal Distributions

Suppose that two independent normal populations are of interest when the population means and variances, say, μ_1,σ_1^2,μ_2,σ_2^2, are unknown. Assume that two random samples of size n_1 from population 1 and of size n_2 from population 2 are available, and let S_1^2 and S_2^2 be the sample variances. We wish to build confidence intervals (F-interval) on the ratio between two population variances

$$\frac{\sigma_1^2}{\sigma_2^2},$$

and test (F-test) the hypotheses

$H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$ (two-tailed)
$H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 < \sigma_2^2$ (left-tailed)
$H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 > \sigma_2^2$ (right-tailed)

Inference on the Variances of Two Normal Distributions

More specifically, we consider two types of inferences on the ratio σ_1^2/σ_2^2 of the variances of two **normal** distributions.

(1) Constructing a $100(1-\alpha)\%$ confidence interval
- Finterval(level=?, sample1=?,sample2=?)
- Finterval(level=?, n1=?,n2=?,s1=?,s2=?)

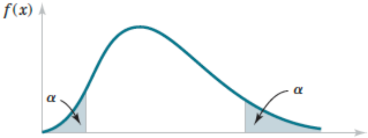
(2) and testing the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against a two-tailed ($\neq$), left-tailed ($<$), or right-tailed ($>$) alternative hypothesis.
- Ftest(H1=?,alpha=?,sample1=?,sample2=?)
- Ftest(H1=?,alpha=?,n1=?,n2=?,s1=?,s2=?)

Inference on the Variances of Two Normal Distributions

The CI starts with a fact that

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \div \frac{\sigma_1^2}{\sigma_2^2} \sim F(n_1 - 1, n_2 - 1),$$

where $F(n_1-1, n_2-1)$ stands for an F-distribution with n_1-1 and n_2-1 degrees of freedom.



Then a $100(1-\alpha)\%$ two-tailed CI for σ_1^2/σ_2^2 comes from

$$f_{n_1-1,n_2-1,1-\alpha/2} \le F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \div \frac{\sigma_1^2}{\sigma_2^2} \le f_{n_1-1,n_2-1,\alpha/2}$$

Inference on the Variances of Two Normal Distributions

When testing $H_0 : \sigma_1^2 = \sigma_2^2$,
The test statistic is $F = \frac{S_1^2}{S_2^2}$.
Based on samples $x_{11}, \dots, x_{1n_1}$ and $x_{21}, \dots, x_{2n_2}$, the observed test statistic is $F_0 = \frac{s_1^2}{s_2^2}$.

- If $H_1 : \sigma_1^2 \neq \sigma_2^2$. At significance level α, we reject H_0 if $F_0 > f_{n_1-1,n_2-1,\alpha/2}$ or $F_0 < f_{n_1-1,n_2-1,1-\alpha/2}$ (critical-value approach),
- If $H_1 : \sigma_1^2 < \sigma_2^2$. At significance level α, we reject H_0 if $F_0 < f_{n_1-1,n_2-1,1-\alpha}$ (critical-value approach),
- If $H_1 : \sigma_1^2 > \sigma_2^2$. At significance level α, we reject H_0 if $F_0 > f_{n_1-1,n_2-1,\alpha}$ (critical-value approach).

We also have the confidence-interval approach and the P-value approach.

Example 6

A company manufactures impellers for use in jet-turbine engines. One of the operations involves grinding a particular surface finish on a titanium alloy component. Two different grinding processes can be used, and both processes can produce parts at identical mean surface roughness. The manufacturing engineer would like to select the process having the least variability in surface roughness. A random sample of $n_1 = 11$ parts from the first process results in a sample standard deviation $s_1 = 5.1$ microinches, and a random sample of $n_2 = 16$ parts from the second process results in a sample standard deviation of $s_2 = 4.7$ microinches. Is there any difference between σ_1 and σ_2 at $\alpha = 0.1$?

Example 6

Solution: We are testing $H_0 : \sigma_1 = \sigma_2$ vs $H_1 : \sigma_1 \neq \sigma_2$ (F-test at $\alpha = 0.1$)

```
Ftest(H1="two",alpha=0.1,n1=11,n2=16,s1=5.1,s2=4.7)
H1 is two-tailed. The results are:

1. Test statistic F0 is 1.177456 ,
f_(n1-1,n2-1,1-alpha/2) is 0.3514918 ,
f_(n1-1,n2-1,alpha/2) is 2.543719 .
Because f_(n1-1,n2-1,1-alpha/2)<=F0<=f_(n1-1,n2-1,alpha/2),
we fail to reject H0 at significance level 0.1

2. A 90% two-tailed confidence interval for the ratio
between two population variances is [ 0.4628876 , 3.349881 ]
which contains 1, so we fail to reject H0 at significance level 0.1

3. The P-value is 0.7508517 which is not smaller than alpha= 0.1 ,
so we fail to reject H0 at significance level 0.1
```

Conclusion: at significance level $\alpha = 0.1$, data do not provide sufficient evidence to reject H_0.

Large-sample Inference on Two Population Proportions

Suppose that two independent random samples of sizes n_1 and n_2 are taken from two populations, and let X_1 and X_2 represent the number of observations that belong to the class of interest in samples 1 and 2, respectively. Denote the population proportions of interest by p_1 and p_2.

We wish to build confidence intervals on the **difference in proportions** ($p_1 - p_2$) and test the hypotheses

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2 \text{ (two-tailed)}$$
$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 < p_2 \text{ (left-tailed)}$$
$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 > p_2 \text{ (right-tailed)}$$

Large-sample Inference on the Variances of Two Normal Distributions

More specifically, we consider two types of inferences on $p_1 - p_2$.

(1) Constructing a $100(1-\alpha)\%$ confidence interval
   - twosample.Propinterval(level=?, n1=?,n2=?,X1=?,X2=?)

(2) and testing the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against a two-tailed ($\neq$), left-tailed ($<$), or right-tailed ($>$) alternative hypothesis.
   - twosample.Proptest(H1=?,alpha=?,n1=?,n2=?,X1=?,X2=?)

Large-sample Inference on the Variances of Two Normal Distributions

We estimate p_1 by $\hat{p}_1 = X_1/n_1$ and p_2 by $\hat{p}_2 = X_2/n_2$. The CI starts with a fact that

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim AN(0,1).$$

Then an **approximate** $100(1-\alpha)\%$ two-tailed CI for $p_1 - p_2$ comes from

$$-z_{\alpha/2} \leq Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \leq z_{\alpha/2}.$$

It is valid if all $n_1\hat{p}_1, n_1(1-\hat{p}_1), n_2\hat{p}_2, n_2(1-\hat{p}_2) \geq 5$.

Large-sample Inference on the Variances of Two Normal Distributions

Require: $n_1\hat{p}_1, n_1(1-\hat{p}_1), n_2\hat{p}_2, n_2(1-\hat{p}_2) \geq 5$
When testing $H_0 : p_1 = p_2$,
The observed test statistic is $z_0 = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$
where $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$.

- If $H_1 : p_1 \neq p_2$. At significance level α, we reject H_0 if $|z_0| > z_{\alpha/2}$ (critical-value approach),
- If $H_1 : p_1 < p_2$. At significance level α, we reject H_0 if $z_0 < -z_\alpha$ (critical-value approach),
- If $H_1 : p_1 > p_2$. At significance level α, we reject H_0 if $z_0 > z_\alpha$ (critical-value approach).

We also have the confidence-interval approach and the P-value approach. The confidence-interval approach **might** give a different conclusion than the other two approaches. But this difference is very minor, especially when sample sizes n_1 and n_2 are large.

Example 7

Extracts of St. John's Wort are widely used to treat depression. An article in the April 18, 2001, issue of the Journal of the American Medical Association ("Effectiveness of St. John's Wort on Major Depression: A Randomized Controlled Trial") compared the efficacy of a standard extract of St. John's Wort with a placebo in 200 outpatients diagnosed with major depression. Patients were randomly assigned to two groups; one group received the St. John's Wort, and the other received the placebo. After eight weeks, 19 of the placebo-treated patients showed improvement, and 27 of those treated with St. John's Wort improved. Is there any reason to believe that St. John's Wort is effective in treating major depression? Use $\alpha = 0.05$.

Example 7

Solution: The parameters of interest are p_1 and p_2, the proportion of patients who improve following treatment with St. John's Wort (p_1) or the placebo (p_2). We are testing $H_0 : p_1 = p_2$ vs $H_1 : p_1 > p_2$ (two-sample proportion Z-test, right-tailed, at $\alpha = 0.05$)

```
twosample.Proptest(H1="right",alpha=0.05,n1=100,n2=100,X1=27,X2=19)
H1 is right-tailed. The results are:

1. Test statistic z0 is 1.344206 , z_alpha is 1.644854 .
Because z0<= z_alpha, we fail to reject H0 at significance level 0.05

2. A 95% one-sided confidence interval for the difference
in population proportions is [ 0.1774498 , Inf )
which contains 0, so we fail to reject H0 at significance level 0.05

3. The P-value is 0.08944095 which is not smaller than alpha= 0.05 ,
so we fail to reject H0 at significance level 0.05
```

Conclusion: at significance level $\alpha = 0.05$, data do not provide sufficient evidence to reject H_0.

StatEngine Summary of Two-Sample Inferences

Z-intervals and Z-tests On $\mu_1 - \mu_2$ (Normality, known σ_1^2 and σ_2^2).

```
twosample.Zinterval(level=?,sigma1=?,sigma2=?,
                    sample1=?,sample2=?)
twosample.Zinterval(level=?,sigma1=?,sigma2=?,
                    barx1=?,barx2=?,n1=?,n2=?)

twosample.Ztest(Delta0=?,H1=?,alpha=?,sigma1=?,sigma2=?,
                sample1=?,sample2=?)
twosample.Ztest(Delta0=?,H1=?,alpha=?,sigma1=?,sigma2=?,
                barx1=?,barx2=?,n1=?,n2=?)
```

StatEngine Summary of Two-Sample Inferences

T-intervals and T-tests On $\mu_1 - \mu_2$ (Normality, unknown σ_1^2 and σ_2^2, pooled=yes or no).

```
twosample.Tinterval(level=?, pooled=?,sample1=?,sample2=?)
twosample.Tinterval(level=?, pooled=?,barx1=?,barx2=?,
                    n1=?,n2=?,s1=?,s2=?)

twosample.Ttest(Delta0=?,H1=?,alpha=?,pooled=yes,
                sample1=?,sample2=?)
twosample.Ttest(Delta0=?,H1=?,alpha=?,pooled=yes,
                barx1=?,barx2=?,n1=?,n2=?,s1=?,s2=?)
```

StatEngine Summary of Two-Sample Inferences

Paired T-intervals and T-tests On $\mu_1 - \mu_2$ (Paires, Normality, unknown varaince σ_D^2).

```
Ttest(mu0=Delta0, H1=?,alpha=?,sample=sample1-sample2)
Ttest(mu0=Delta0, H1=?,alpha=?,n=?,
      barx=mean(sample1-sample2),s=sd(sample1-sample2))

Tinterval(level=?,sample=sample1-sample2)
Tinterval(level=?,n=?,
      barx=mean(sample1-sample2),s=sd(sample1-sample2))
```

F-intervals and F-tests On σ_1^2/σ_2^2 (Normality).

```
Finterval(level=?, sample1=?,sample2=?)
Finterval(level=?, n1=?,n2=?,s1=?,s2=?)

Ftest(H1=?,alpha=?,sample1=?,sample2=?)
Ftest(H1=?,alpha=?,n1=?,n2=?,s1=?,s2=?)
```

Two-Sample Proportion Z-intervals and Z-tests On $p_1 - p_2$ ($n_1\hat{p}_1, n_1(1-\hat{p}_1), n_2\hat{p}_2, n_2(1-\hat{p}_2) \geq 5$).

```
twosample.Propinterval(level=?, n1=?,n2=?,X1=?,X2=?)

twosample.Proptest(H1=?,alpha=?,n1=?,n2=?,X1=?,X2=?)
```