Lecture 13.1, MATH-57091 Probability and Statistics for High-School Teachers.

Artem Zvavitch

Department of Mathematical Sciences, Kent State University

November 26, 2014.

Example

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.

Example

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table below cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Tire Lives in units of 1000 kilometers

Tires tested at A	Tires tested at B
66.4	58.2
61.6	60.4
60.5	55.2
59.1	62.0
63.6	57.3
61.4	58.7
62.5	56.1
64.4	
60.7	

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \dots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \dots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \dots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2.

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \dots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \dots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2. We are interested to test a null hypothesis that two population means are equal:

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \ldots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2. We are interested to test a null hypothesis that two population means are equal:

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

It is logical, as before to use respective sample means to estimate μ_X and μ_Y:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \ldots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2. We are interested to test a null hypothesis that two population means are equal:

$$H_0 : \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1 : \quad \mu_X \neq \mu_Y .$$

It is logical, as before to use respective sample means to estimate μ_X and μ_Y:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i$$

and thus to reject H_0 if $\bar{X}$ and $\bar{Y}$ are far from each other.

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \ldots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2. We are interested to test a null hypothesis that two population means are equal:

$$H_0 : \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1 : \quad \mu_X \neq \mu_Y.$$

It is logical, as before to use respective sample means to estimate μ_X and μ_Y:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

and thus to reject H_0 if $\bar{X}$ and $\bar{Y}$ are far from each other. To measure "far" we again need to find an appropriate constant c such that

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X} - \bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

How do we find c?

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \ldots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2. We are interested to test a null hypothesis that two population means are equal:

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

It is logical, as before to use respective sample means to estimate μ_X and μ_Y:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

and thus to reject H_0 if $\bar{X}$ and $\bar{Y}$ are far from each other. To measure "far" we again need to find an appropriate constant c such that

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X} - \bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

How do we find c? As before we need to determine how large c needs to be to justify rejection at the level of significance α (i.e. probability of error of the first type - to reject something which is actually true).

Testing Equality of Means of two normal populations: case of known variance.

Suppose that $X_1, \dots, X_n$ are sample from normal population having **unknown** mean μ_X and known variance σ_X^2 and suppose that $Y_1, \dots, Y_m$ are sample from normal population having **unknown** mean μ_Y and known variance σ_Y^2. We are interested to test a null hypothesis that two population means are equal:

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

It is logical, as before to use respective sample means to estimate μ_X and μ_Y:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

and thus to reject H_0 if $\bar{X}$ and $\bar{Y}$ are far from each other. To measure "far" we again need to find an appropriate constant c such that

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X} - \bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

How do we find c? As before we need to determine how large c needs to be to justify rejection at the level of significance α (i.e. probability of error of the first type - to reject something which is actually true). For this, we need to know the probability distribution of $\bar{X} - \bar{Y}$.

Probability distribution of $\bar{X} - \bar{Y}$.

Suppose that $X_1, \dots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \dots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

Probability distribution of $\bar{X} - \bar{Y}$.

Suppose that $X_1, \dots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \dots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Suppose that $X_1, \ldots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \ldots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Then $\bar{X}$ is normal with mean μ_X and variance σ_X^2/n and $\bar{Y}$ is normal with mean μ_Y and variance σ_Y^2/m,

Probability distribution of $\bar{X} - \bar{Y}$.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \ldots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Then $\bar{X}$ is normal with mean μ_X and variance σ_X^2/n and $\bar{Y}$ is normal with mean μ_Y and variance σ_Y^2/m, thus $\bar{X} - \bar{Y}$ is also normal with mean

$$\mathbb{E}(\bar{X} - \bar{Y}) = \mathbb{E}\bar{X} - \mathbb{E}\bar{Y} = \mu_X - \mu_Y$$

Probability distribution of $\bar{X} - \bar{Y}$.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \ldots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Then $\bar{X}$ is normal with mean μ_X and variance σ_X^2/n and $\bar{Y}$ is normal with mean μ_Y and variance σ_Y^2/m, thus $\bar{X} - \bar{Y}$ is also normal with mean

$$\mathbb{E}(\bar{X} - \bar{Y}) = \mathbb{E}\bar{X} - \mathbb{E}\bar{Y} = \mu_X - \mu_Y$$

and variance

$$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(-\bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \sigma_X^2/n + \sigma_Y^2/m.$$

Probability distribution of $\bar{X} - \bar{Y}$.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \ldots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Then $\bar{X}$ is normal with mean μ_X and variance σ_X^2/n and $\bar{Y}$ is normal with mean μ_Y and variance σ_Y^2/m, thus $\bar{X} - \bar{Y}$ is also normal with mean

$$\mathbb{E}(\bar{X} - \bar{Y}) = \mathbb{E}\bar{X} - \mathbb{E}\bar{Y} = \mu_X - \mu_Y$$

and variance

$$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(-\bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \sigma_X^2/n + \sigma_Y^2/m.$$

Hence

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

is our usual $N(0,1)$ random variable (standard normal distribution).

Probability distribution of $\bar{X} - \bar{Y}$.

Suppose that $X_1, \ldots, X_n$ are sample from normal population having mean μ_X and variance σ_X^2 and that $Y_1, \ldots, Y_m$ are sample from normal population having mean μ_Y and variance σ_Y^2. Moreover, assume that those are independent samples!

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Then $\bar{X}$ is normal with mean μ_X and variance σ_X^2/n and $\bar{Y}$ is normal with mean μ_Y and variance σ_Y^2/m, thus $\bar{X} - \bar{Y}$ is also normal with mean

$$\mathbb{E}(\bar{X} - \bar{Y}) = \mathbb{E}\bar{X} - \mathbb{E}\bar{Y} = \mu_X - \mu_Y$$

and variance

$$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(-\bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \sigma_X^2/n + \sigma_Y^2/m.$$

Hence

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

is our usual $N(0,1)$ random variable (standard normal distribution). Thus, when H_0 is true ($\mu_X - \mu_Y = 0$) the test statistics TS given by

$$TS = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

will be a standard normal distribution!

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X} - \bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c!

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X} - \bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c! But now we know that

$$TS = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

will be a standard normal distribution.

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X}-\bar{Y}|>c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c! But now we know that

$$TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$$

will be a standard normal distribution. For standard normal variable Z we use notation

$$P(|Z| \geq z_{\alpha/2}) = 2P(Z \geq z_{\alpha/2}) = \alpha.$$

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X}-\bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c! But now we know that

$$TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$$

will be a standard normal distribution. For standard normal variable Z we use notation

$$P(|Z| \geq z_{\alpha/2}) = 2P(Z \geq z_{\alpha/2}) = \alpha.$$

Finally, the appropriate significance-level-α test of

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

is to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |TS| > z_{\alpha/2}, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X}-\bar{Y}| > c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c! But now we know that

$$TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$$

will be a standard normal distribution. For standard normal variable Z we use notation

$$P(|Z| \geq z_{\alpha/2}) = 2P(Z \geq z_{\alpha/2}) = \alpha.$$

Finally, the appropriate significance-level-α test of

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

is to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |TS| > z_{\alpha/2}, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

An alternative way of carrying out this test is first to compute $TS = \nu$.

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |\bar{X}-\bar{Y}|>c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c! But now we know that

$$TS=\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$$

will be a standard normal distribution. For standard normal variable Z we use notation

$$P(|Z|\geq z_{\alpha/2})=2P(Z\geq z_{\alpha/2})=\alpha.$$

Finally, the appropriate significance-level-α test of

$$H_0: \quad \mu_X=\mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X\neq\mu_Y.$$

is to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |TS|>z_{\alpha/2}, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

An alternative way of carrying out this test is first to compute $TS=\nu$. The resulting p value for the test of H_0 versus H_1 is the probability that the absolute value of a standard normal variable is at least as large as $|\nu|$:

$$p \text{ value } = P(|Z|\geq|\nu|)=2P(Z\geq|\nu|).$$

Testing Equality of Means of two normal populations: case of known Variance.

We decided to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } |\bar{X}-\bar{Y}|>c, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

Our goal is to find c! But now we know that

$$TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$$

will be a standard normal distribution. For standard normal variable Z we use notation

$$P(|Z| \geq z_{\alpha/2}) = 2P(Z \geq z_{\alpha/2}) = \alpha.$$

Finally, the appropriate significance-level-α test of

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } |TS|>z_{\alpha/2}, \\ \text{Not Reject } H_0 & \text{otherwise.} \end{array}$$

An alternative way of carrying out this test is first to compute $TS=\nu$. The resulting p value for the test of H_0 versus H_1 is the probability that the absolute value of a standard normal variable is at least as large as $|\nu|$:

$$p \text{ value } = P(|Z| \geq |\nu|) = 2P(Z \geq |\nu|).$$

After we found the p value we reject H_0 is p is small enough (for example, if we are given α we will reject H_0 if $p \leq \alpha$).

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table below cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Tire Lives in units of 1000 kilometers

Tires tested at A	Tires tested at B
66.4	58.2
61.6	60.4
60.5	55.2
59.1	62.0
63.6	57.3
61.4	58.7
62.5	56.1
64.4	
60.7	

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table below cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Tire Lives in units of 1000 kilometers

Tires tested at A	Tires tested at B
66.4	58.2
61.6	60.4
60.5	55.2
59.1	62.0
63.6	57.3
61.4	58.7
62.5	56.1
64.4	
60.7	

Solution: So we will call the tires tested at A the X sample and at B the Y sample and test

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table below cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Tire Lives in units of 1000 kilometers

Tires tested at A	Tires tested at B
66.4	58.2
61.6	60.4
60.5	55.2
59.1	62.0
63.6	57.3
61.4	58.7
62.5	56.1
64.4	
60.7	

Solution: So we will call the tires tested at A the X sample and at B the Y sample and test

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table below cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Tire Lives in units of 1000 kilometers

Tires tested at A	Tires tested at B
66.4	58.2
61.6	60.4
60.5	55.2
59.1	62.0
63.6	57.3
61.4	58.7
62.5	56.1
64.4	
60.7	

Solution: So we will call the tires tested at A the X sample and at B the Y sample and test

$$H_0 : \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1 : \quad \mu_X \neq \mu_Y .$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{Y} = 58.27$.

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table (previous slide) cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Solution: So we will call the tires tested at A the X sample and at B the Y sample so we test

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{Y} = 58.27$. Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_Y = 4$ we get:

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table (previous slide) cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Solution: So we will call the tires tested at A the X sample and at B the Y sample so we test

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{Y} = 58.27$. Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_Y = 4$ we get:

$$TS = \frac{62.24 - 58.27}{\sqrt{9/9 + 16/7}} = 2.19$$

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table (previous slide) cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Solution: So we will call the tires tested at A the X sample and at B the Y sample so we test

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{Y} = 58.27$. Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_Y = 4$ we get:

$$TS = \frac{62.24 - 58.27}{\sqrt{9/9 + 16/7}} = 2.19$$

Thus the p value is equal to (use tables for normal distribution):

$$p \text{ value } = P(|Z| \geq 2.19) = 2P(Z \geq 2.19) = 0.028$$

Back to our example:

Two new methods for producing a tire have been proposed. The manufacturer believes there will be no appreciable difference in the lifetimes of tires produced by the methods. To test the plausibility of such a hypothesis, a sample of 9 tires is produced by method 1 and a sample of 7 tires by method 2. The first sample of tires is to be road tested at a location A and the second at location B.
It is known from previous experience that the lifetime of those tires are normal random variable with a mean life due to the tire but variance due to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with SD equal to 3000 kilometers, and at location B are normal with SD 4000 kilometers.
Should the data at the table (previous slide) cause the manufacturer to reject the hypothesis that the mean lifetime is the same for both types? Use 5 percent level; of significance.

Solution: So we will call the tires tested at A the X sample and at B the Y sample so we test

$$H_0: \quad \mu_X = \mu_Y \quad \text{against the alternative} \quad H_1: \quad \mu_X \neq \mu_Y.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{Y} = 58.27$. Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_Y = 4$ we get:

$$TS = \frac{62.24 - 58.27}{\sqrt{9/9 + 16/7}} = 2.19$$

Thus the p value is equal to (use tables for normal distribution):

$$p \text{ value } = P(|Z| \geq 2.19) = 2P(Z \geq 2.19) = 0.028$$

the hypothesis of equal means is rejected at any significance level greater or equal then 2.8 percent, in particular we reject it at 5 percent level.

One-sided tests

Absolutely the same ideology should be used for one sided test!

One-sided tests

Absolutely the same ideology should be used for one sided test! Before we were studying hypothesis of type $\mu_X = \mu_Y$.

One-sided tests

Absolutely the same ideology should be used for one sided test! Before we were studying hypothesis of type $\mu_X = \mu_Y$. Quite often we are interested to test

$$H_0: \quad \mu_X \le \mu_Y \text{ against}$$
$$H_1: \quad \mu_X > \mu_Y$$

One-sided tests

Absolutely the same ideology should be used for one sided test! Before we were studying hypothesis of type $\mu_X = \mu_Y$. Quite often we are interested to test

$$\begin{array}{ll} H_0: & \mu_X \le \mu_Y \text{ against} \\ H_1: & \mu_X > \mu_Y \end{array}$$

We will want to reject H_0 when TS (test statistics) is large, it can be shown in exactly the same way as in two-sided case that the significance-level-α test is to

$$\begin{array}{ll} \text{Reject } H_0: & TS \ge z_\alpha \\ \text{Not Reject } H_0: & TS < z_\alpha. \end{array}$$

where

$$TS = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

Absolutely the same ideology should be used for one sided test! Before we were studying hypothesis of type $\mu_X = \mu_Y$. Quite often we are interested to test

$$\begin{aligned} H_0: &\quad \mu_X \le \mu_Y \text{ against} \\ H_1: &\quad \mu_X > \mu_Y \end{aligned}$$

We will want to reject H_0 when TS (test statistics) is large, it can be shown in exactly the same way as in two-sided case that the significance-level-α test is to

$$\begin{aligned} \text{Reject } H_0: &\quad TS \ge z_\alpha \\ \text{Not Reject } H_0: &\quad TS < z_\alpha. \end{aligned}$$

where

$$TS = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

Equivalently, if the observed value of TS is ν, then the p value us

$$p \text{ value} = P(Z \ge \nu),$$

Absolutely the same ideology should be used for one sided test! Before we were studying hypothesis of type $\mu_X = \mu_Y$. Quite often we are interested to test

$$\begin{array}{ll} H_0: & \mu_X \leq \mu_Y \text{ against} \\ H_1: & \mu_X > \mu_Y \end{array}$$

We will want to reject H_0 when TS (test statistics) is large, it can be shown in exactly the same way as in two-sided case that the significance-level-α test is to

$$\begin{array}{ll} \text{Reject } H_0: & TS \geq z_\alpha \\ \text{Not Reject } H_0: & TS < z_\alpha. \end{array}$$

where

$$TS = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}$$

Equivalently, if the observed value of TS is ν, then the p value us

$$p \text{ value} = P(Z \geq \nu),$$

The null hypothesis is then rejected at any significance level greater than or equal to the value p.

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$.

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) .

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) . So we should test

$$H_0: \quad \mu_X \le \mu_W \quad \text{against the alternative} \quad H_1: \quad \mu_X > \mu_W.$$

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) . So we should test

$$H_0 : \quad \mu_X \le \mu_W \quad \text{against the alternative} \quad H_1 : \quad \mu_X > \mu_W .$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{W} = 59.27$ (remember $\bar{W} = \bar{Y} + 1$).

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) . So we should test

$$H_0 : \quad \mu_X \leq \mu_W \quad \text{against the alternative} \quad H_1 : \quad \mu_X > \mu_W.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{W} = 59.27$ (remember $\bar{W} = \bar{Y} + 1$). Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_W = 4$ (again we use that $\sigma_W = \sigma_Y$, i.e. SD does not change when we shift all the data by the same number) we get:

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) . So we should test

$$H_0: \quad \mu_X \le \mu_W \quad \text{against the alternative} \quad H_1: \quad \mu_X > \mu_W.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{W} = 59.27$ (remember $\bar{W} = \bar{Y} + 1$). Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_W = 4$ (again we use that $\sigma_W = \sigma_Y$, i.e. SD does not change when we shift all the data by the same number) we get:

$$TS = \frac{62.24 - 59.27}{\sqrt{9/9 + 16/7}} = 1.64.$$

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \ldots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) . So we should test

$$H_0 : \quad \mu_X \leq \mu_W \quad \text{against the alternative} \quad H_1 : \quad \mu_X > \mu_W.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{W} = 59.27$ (remember $\bar{W} = \bar{Y} + 1$). Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_W = 4$ (again we use that $\sigma_W = \sigma_Y$, i.e. SD does not change when we shift all the data by the same number) we get:

$$TS = \frac{62.24 - 59.27}{\sqrt{9/9 + 16/7}} = 1.64.$$

Thus the p value is equal to (use tables for normal distribution):

$$p \text{ value } = P(Z \geq 1.64) = 0.0505.$$

Example:

Suppose the purpose of the experiment in our previous example about tire company was to attempt to prove the hypothesis that the mean life of the first set of tires exceeded that of the second set by more than 1000 kilometers. Are the data strong enough to establish that at 5 percent level of significance?

Solution: As before we call the tires tested at A the X sample and at B the Y sample. We will make an additional small trick before starting the calculations.
Let $W_i = Y_i + 1$ for $i = 1, 2, \dots, 7$, then we are interested in determining whether the data will enable us to conclude $\mu_X > \mu_W$. To decide this we need "to reject" the alternative hypothesis (remember we never "accept" anything!) . So we should test

$$H_0 : \quad \mu_X \leq \mu_W \quad \text{against the alternative} \quad H_1 : \quad \mu_X > \mu_W.$$

To compute test statistics TS we first evaluate $\bar{X} = 62.24$ and $\bar{W} = 59.27$ (remember $\bar{W} = \bar{Y} + 1$). Since $n = 9$, $m = 7$, $\sigma_X = 3$ and $\sigma_W = 4$ (again we use that $\sigma_W = \sigma_Y$, i.e. SD does not change when we shift all the data by the same number) we get:

$$TS = \frac{62.24 - 59.27}{\sqrt{9/9 + 16/7}} = 1.64.$$

Thus the p value is equal to (use tables for normal distribution):

$$p \text{ value } = P(Z \geq 1.64) = 0.0505.$$

So p value is greater then 5 percent, and thus thought the evidence is strongly in favor of the alternative hypothesis it is not quite strong enough to let us to reject the H_0 at the 5 percent level of significance.

Summary for the Hypothesis test concerning the mean of two Normal populations with known variance and when samples are independent.

Summary for the Hypothesis test concerning the mean of two Normal populations with known variance and when samples are independent.

Let $X_1, \dots, X_n$ are a sample of size n from a normal population with **unknown** mean μ_X, **known SD** σ_X and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Summary for the Hypothesis test concerning the mean of two Normal populations with known variance and when samples are independent.

Let $X_1, \dots, X_n$ are a sample of size n from a normal population with **unknown** mean μ_X, **known SD** σ_X and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $Y_1, \dots, Y_m$ are a sample of size m from a normal population with **unknown** mean μ_Y, **known SD** σ_Y and $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$. Finally, we assume that two samples are independent! Then

Summary for the Hypothesis test concerning the mean of two Normal populations with known variance and when samples are independent.

Let $X_1, \dots, X_n$ are a sample of size n from a normal population with **unknown** mean μ_X, **known SD** σ_X and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $Y_1, \dots, Y_m$ are a sample of size m from a normal population with **unknown** mean μ_Y, **known SD** σ_Y and $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$. Finally, we assume that two samples are independent! Then

H_0 :	H_1	Test Statistics (TS)	Significance-level-α test	p Value if TS$=\nu$
$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$	Reject H_0 if $\lvert TS\rvert \geq z_{\alpha/2}$. Do not reject H_0 otherwise	$2P(Z \geq \lvert\nu\rvert)$
$\mu_Z \leq \mu_Y$	$\mu_X > \mu_Y$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}$	Reject H_0 if $TS \geq z_\alpha$. Do not reject H_0 otherwise	$P(Z \geq \nu)$