10A 2SamZ-test

Contents

Textbook: Devore 8e

A Two-Sample Z-test

[ToC]

A.1 Add or Subtract Two Normal Random Variables

Suppose

$$X \sim N(\mu_1, \sigma_1) \qquad \text{and} \qquad Y \sim N(\mu_2, \sigma_2),$$

X and Y are independent. What are the distribution of

$$X + Y \sim ?$$
$$X - Y \sim ?$$

A.2 Variance Adds, NOT Std Dev

Since any linear combination of two normals are also normal, we have

$$X + Y \sim N(\mu_1 + \mu_2, \quad \sigma_1^2 + \sigma_2^2)$$

And also

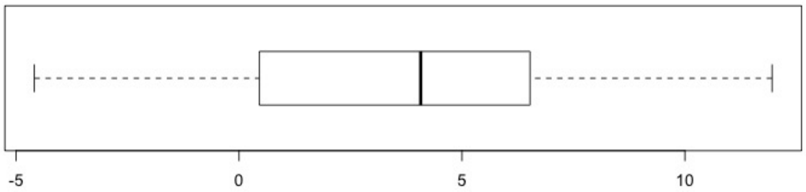
$$X - Y \sim N(\mu_1 - \mu_2, \quad \sigma_1^2 + \sigma_2^2)$$

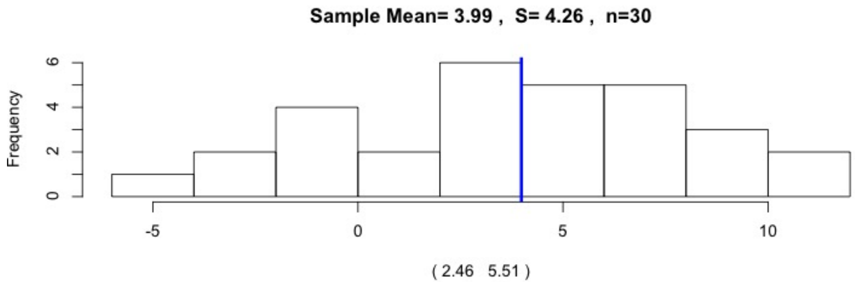
A.3 One-Sample Inference Review

One-sample inference: There's population with true mean μ. Want to know μ.

- Estimate μ $\Rightarrow$ CI
- Is $\mu > \mu_0$? $\Rightarrow$ See if μ_0 is in CI

  Test of hypothesis (z-test, t-test)


-5
0
5
10

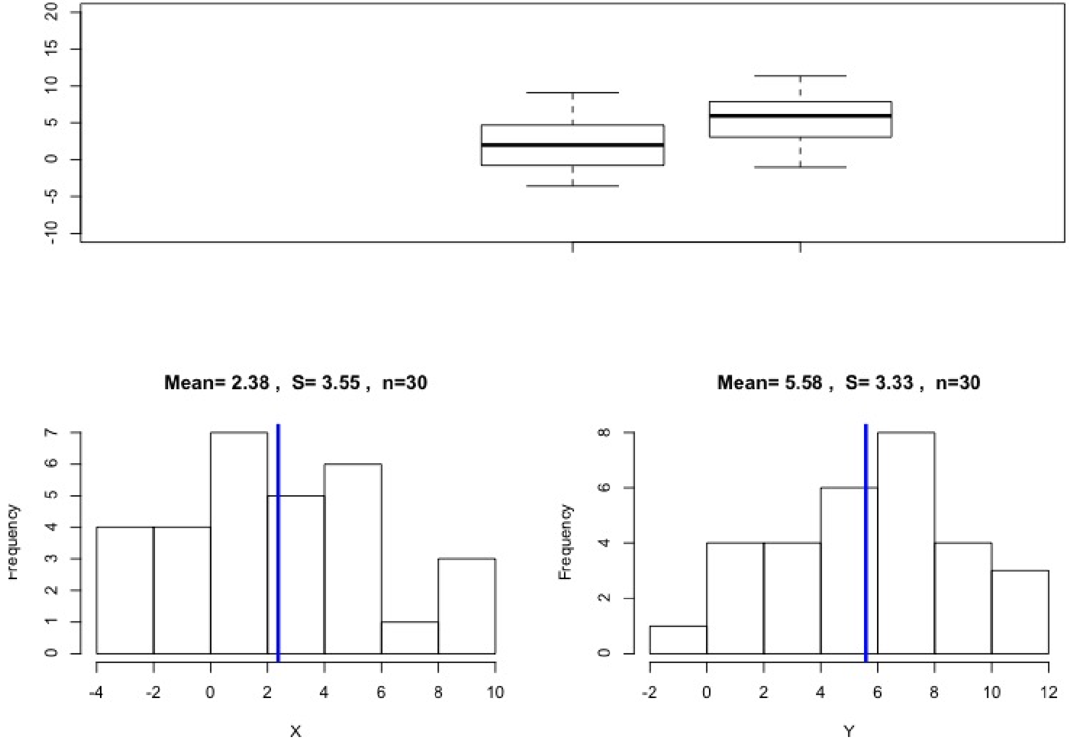

Sample Mean= 3.99 , S= 4.26 , n=30
Frequency
0
2
4
6
-5
0
5
10
(2.46   5.51)

A.4 Two-sample inference:

Let's use the same principle, and get inference on two-sample problem.
There's population 1 with true mean μ_1.
There's population 2 with true mean μ_2.

- Can we estimate the difference $\mu_1 - \mu_2$? $\Rightarrow$ CI

- Is $\mu_1 = \mu_2$? Is $\mu_1 > \mu_2 + 10$? μ $\Rightarrow$ Test of hypothesis



Mean= 2.38 , S= 3.55 , n=30
Mean= 5.58 , S= 3.33 , n=30
Frequency
Frequency
X
Y

A.5 Two-sample notations

- Assume population 1 has $N(\mu_1, \sigma_1^2)$ distribution. Then draw sample $X_1, \ldots, X_{n_1}$ from population 1.

  Calculate $\bar{X}$, S_1. Sample size $= n_1$.

- Assume population 2 has $N(\mu_2, \sigma_2^2)$ distribution. Then draw sample $Y_1, \ldots, Y_{n_2}$ from population 2.

  Calculate $\bar{Y}$, S_2. Sample size $= n_2$.

A.6 Two Estimators

We know that

$$\bar{X} \sim N\Big(\mu_1, \frac{\sigma_1^2}{n_1}\Big) \qquad \text{and} \qquad \bar{Y} \sim N\Big(\mu_2, \frac{\sigma_2^2}{n_2}\Big).$$

Then, since they are independent,

$$\bar{X} - \bar{Y} \sim N\Big(\mu_1 - \mu_2, \quad \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\Big)$$

This looks kinda like in one-sample case,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

A.7 Use Analogy to write CI for $\mu_1 - \mu_2$

one-sample case	two-sample case
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$	$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \quad \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}\right)$
CI for μ	CI for $\mu_1 - \mu_2$
$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$(\overline{X} - \overline{Y}) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

A.8 Use Analogy to write test for $\mu_1 - \mu_2$

one-sample case	two-sample case
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$	$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \quad \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}\right)$
$H_0 : \mu = \mu_0$	$H_0 : \mu_1 = \mu_2$
$H_A : \mu > \mu_0$	$H_A : \mu_1 > \mu_2$
$\mu < \mu_0$	$\mu_1 < \mu_2$
$\mu \neq \mu_0$	$\mu_1 \neq \mu_2$

We need to rewrite the hypothesis

A.9 Re-write the hypothesis

one-sample case	two-sample case	if $\Delta_0 = 0$
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$	$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \quad \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}\right)$	
$H_0 : \mu = \mu_0$	$H_0 : \mu_1 - \mu_2 = \Delta_0$	$(\mu_1 = \mu_2)$
$H_A : \mu > \mu_0$	$H_A : \mu_1 - \mu_2 = \Delta_0$	$(\mu_1 > \mu_2)$
$\mu < \mu_0$	$\mu_1 - \mu_2 < \Delta_0$	$(\mu_1 < \mu_2)$
$\mu \neq \mu_0$	$\mu_1 - \mu_2 \neq \Delta_0$	$(\mu_1 \neq \mu_2)$
$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	

A.10 Two-sample z-test

To test the null hypothesis of

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

against alternatives

$$\begin{aligned} H_A : \mu_1 - \mu_2 &> \Delta_0 \quad \text{(Upper-tailed alternative)} \\ \mu_1 - \mu_2 &< \Delta_0 \quad \text{(Lower-tailed alternative)} \\ \mu_1 - \mu_2 &\neq \Delta_0 \quad \text{(Two-tailed alternative)}\ , \end{aligned}$$

we use the test statistic

$$z = \frac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

and rest is the same old z-test:

H_A	rejection region	p-value	power
upper-tailed	$z > z_\alpha$	$1 - \Phi(z)$	$1 - \Phi(z_\alpha - \mu_A^*)$
lower-tailed	$z < -z_\alpha$	$\Phi(z)$	$\Phi(-z_\alpha - \mu_A^*)$
Two-tailed	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	$2\big(1 - \Phi(\lvert z \rvert)\big)$	$1 - \Phi(z_{\frac{\alpha}{2}} - \mu_A^*) + \Phi(-z_{\frac{\alpha}{2}} - \mu_A^*)$

$$\mu_A^* = \frac{\mu_1 - \mu_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

A.11 Ex: Tire and Fuel

A taxi company manager is trying to decide whether the use of radial tires instead of regular belted tires improves fuel economy. Twelve cars were equipped with radial tires and driven over a prescribed test course. Without changing drivers, the same cars were then equipped with regular belted tires and driven once again over the test course. The gasoline consumption, in kilometers per liter, was recorded as follows:

Can we conclude that cars equipped with radial tires give better fuel economy than those equipped with belted tires? Assume the populations to be normally distributed.

Assume that the observations are random sample from the Normal distributions, and $\sigma_1 = \sigma_2 = 1$

A.12 Data: (tire and fuel)

```
    Car Radial Belted  (Km/l)
D=c(1 ,   4.2,    4.1,
    2 ,   4.7,    4.4,
    3 ,   6.6,    6.4,
    4 ,   7.0,    6.7,
    5 ,   6.7,    6.4,
    6 ,   4.5,    4.4,
    7 ,   5.7,    5.7,
    8 ,   6.0,    5.8,
    9 ,   7.4,    6.5,
    10,   4.9,    4.7,
    11,   6.1,    6.0,
    12,   5.2,    4.9)

dim(D) <- c(3,12)
D <- t(D)

t.test(D[,2], D[,3])
```

$\bar{X} = 5.75 \quad \bar{Y} = 5.50$ Is this an evidence that $\mu_1 > \mu_2$?

$S_1^2 = 1.053 \quad S_2^2 = 0.945$

Assuming $\sigma_1^2 = \sigma_2^2 = 1$, we want to test

$$H_0: \quad \mu_1 - \mu_2 = 0$$

$$H_A: \quad \mu_1 - \mu_2 > 0 \quad (upper - tailed)$$

Test Statistic

$$z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5.75 - 5.50 - 0}{\sqrt{\frac{1}{12} + \frac{1}{12}}} = .6124$$

P-value $= 1 - \Phi(.6124) = .27.$