Hypothesis testing

Anastasiia Kim

May 4, 2020

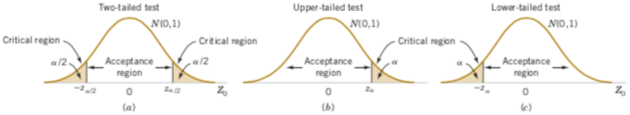
Z-test for a population mean. Normal distribution. Variance σ^2 is known

Testing Hypotheses on the Mean, Variance Known (Z-Tests)

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

Alternative Hypotheses	P-Value	Rejection Criterion for Fixed-Level Tests
H_1: $\mu \neq \mu_0$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$. $P = 2\left[1 - \Phi(\lvert z_0 \rvert)\right]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
H_1: $\mu > \mu_0$	Probability above z_0. $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
H_1: $\mu < \mu_0$	Probability below z_0. $P = \Phi(z_0)$	$z_0 < -z_\alpha$



Example

Air crew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic and follows Normal distribution. Specifications require that the mean burning rate must be 50 centimeters per second. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a *type I error probability or significance level* of $\alpha = 0.05$ and selects a random sample of n $= 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusions should be drawn?

- The parameter of interest is μ, the mean burning rate
- Hypotheses: $H_0 : \mu = 50$ and $H_1 : \mu \neq 50$
- Given that the distribution is Normal with known σ, the test statistic is

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{25}} = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

Example

- The parameter of interest is μ, the mean burning rate
- Hypotheses: $H_0 : \mu = 50$ and $H_1 : \mu \neq 50$
- Given that the distribution is Normal with known σ, the test statistic is

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{25}} = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

- Two ways to decide whether we should reject H_0:
  - Use the value of test statistic to reject H_0 if

$$\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}$$

$$\text{compare } |3.25| \text{ with } z_{0.05/2} = 1.96$$

    $3.25 > 1.96$ reject H_0.
  - Or Use p-value and reject H_0 if the p-value is less than α. Since H_1 is two-sided ($\neq$) then p-value $= 2(1 - \Phi(3.25)) = 0.0012$ which is less than $\alpha = 0.05$.
  - Draw conclusion: since we reject H_0 in favor H_1, the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

Type I and Type II errors

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

Power of the test

The power of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true. The power is computed as $1-\beta$, and power can be interpreted as the probability of correctly rejecting a false null hypothesis.

- The parameter of interest is μ
- Hypotheses: $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$
- Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, then the probability of a type II Error for a two-sided test

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

  where $\Phi(z)$ denotes the probability to the left of z in the standard normal distribution.
- Power is $1-\beta$, a concise measure of the sensitivity of a statistical test when by sensitivity we mean the ability of the test to detect differences.

Power of the test

- The parameter of interest is μ, the mean burning rate
- Hypotheses: $H_0 : \mu = 50$ and $H_1 : \mu \neq 50$
- Suppose that the true value of the mean is $\mu = 52$ so $\delta = \mu - \mu_0 = 52 - 50 = 2$
- When $n = 10$, $\sigma = 2$, and $\alpha = 0.05$, we can find that β as

$$\beta = \Phi\left(z_{0.05/2} - \frac{2\sqrt{10}}{2}\right) - \Phi\left(-z_{0.05/2} - \frac{2\sqrt{10}}{2}\right) =$$

$$\Phi(-1.202) - \Phi(-5.122) = 0.1146 - 1.509335e - 07 = 0.1146$$

The power is $1 - \beta = 1 - 0.1146 = 0.8854$

Therefore, the sensitivity of the test for detecting the difference between a mean burning rate of 50 centimeters per second and 52 centimeters per second is 0.8854. That is, if the true mean is really 52 centimeters per second, this test will correctly reject $H_0 : \mu = 50$ and 'detect' this difference 88.54% of the time. If this value of power is judged to be too low, we can increase either the significance level α or the sample size n.

Power of the test. Sample size

Determine the sample size to obtain a particular value of β, and thus power $1-\beta$ for a given δ and α.

- $\delta = \mu - \mu_0$
- sample size for two-sided test $(H_1 : \mu \neq \mu_0)$ on the mean, variance known:

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2}$$

- sample size for one-sided test $(H_1 : \mu > \mu_0, \mu < \mu_0)$ on the mean, variance known:

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\delta^2}$$

- round up is n is not integer.

Power of the test. Sample size. R

$$H_0 : \mu = 50 \qquad H_1 : \mu \neq 50$$

Suppose that the true burning rate is 49 centimeters per second. Determine n to achieve the power of 0.90 for the two-sided test with $\alpha = 0.05$, $\sigma = 2$, and $\delta = 49 - 50 = -1$.

$$n = \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2} = \frac{(z_{0.05/2} + z_{0.10})^2 2^2}{(-1)^2} = (1.96 + 1.28)^2 4 = 41.99$$

To achieve a power of 0.90, we need a sample size of $n = 42$.
R: $z_{0.05/2}$: $qnorm(1 - 0.05/2) = 1.96$ and $z_{0.10}$: $qnorm(1 - 0.10) = 1.28$. In R, $qnorm(y)$ outputs the value of inverse cdf $F^{-1}(y)$. If $y = F(x)$ then $x = F^{-1}(y)$. That is qnorm looks up the y-th quantile of the standard normal distribution. What is the z-score of the 99th quantile of the distribution? $qnorm(0.99) = 2.326$ which is $z_{0.01}$. Also $pnorm(2.326) = 0.99$ gives cdf $F(2.326)$.

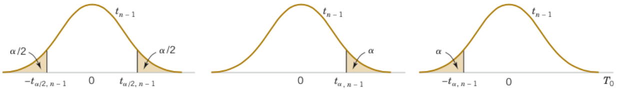
T-test for a population mean

Testing Hypotheses on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $T_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$

Alternative Hypotheses	***P*-Value**	**Rejection Criterion for Fixed-Level Tests**
$H_1: \mu \neq \mu_0$	Probability above $\lvert t_0 \rvert$ and probability below $-\lvert t_0 \rvert$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1: \mu > \mu_0$	Probability above t_0	$t_0 > t_{\alpha,n-1}$
$H_1: \mu < \mu_0$	Probability below t_0	$t_0 < -t_{\alpha,n-1}$



(a) (b) (c)

Example

- The parameter of interest is μ, the mean
- Hypotheses: $H_0 : \mu = 0.82$ and $H_1 : \mu > 0.82$
- Suppose that $\bar{x} = 0.83725$, $n = 15$, and $s = 0.02456$, $\alpha = 0.05$
- Given that the distribution is Normal with unknown σ, the test statistic is

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

- Reject H_0 (one-sided) if the absolute value of test statistic $|2.72| > t_{\alpha,n-1}$

$$t_{0.05,15-1} = 1.76 \qquad R : qt(1 - 0.05, 15 - 1)$$

we reject H_0 and conclude that the mean exceeds 0.82.

Example

- The parameter of interest is μ, the mean
- Hypotheses: $H_0 : \mu = 0.82$ and $H_1 : \mu > 0.82$
- Suppose that $\bar{x} = 0.83725$, $n = 15$, and $s = 0.02456$, $\alpha = 0.05$

Find the power of the test given that the true mean is 0.84 so $\delta = 0.84 - 0.82 = 0.02$:

```
power.t.test(n = 15, delta = 0.02, sd=0.02456, type="one.sample")

     One-sample t test power calculation

              n = 15
          delta = 0.02
             sd = 0.02456
      sig.level = 0.05
          power = 0.8344035
    alternative = two.sided

power.t.test(n = 15, delta = 0.02, sd=0.02456, type="one.sample", alternative="one.sided")

     One-sample t test power calculation

              n = 15
          delta = 0.02
             sd = 0.02456
      sig.level = 0.05
          power = 0.911696
    alternative = one.sided
```

The power is 0.91. Thus, the probability of rejecting $H_0 : \mu = 0.82$ if the true mean exceeds this by 0.02 is approximately $1 - \beta = 0.91$.