STAT 509: Statistics for Engineers

Chapter 9: Tests of Hypotheses for a Single Sample

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

Fall 2020

Learning Objectives:

1. Structure engineering decision-making problems as hypothesis tests
2. Test hypotheses on the mean of a normal distribution using either a Z-test or a t-test procedure
3. Test hypotheses on the variance or standard deviation of a normal distribution
4. Test hypotheses on a population proportion
5. Use the P-value approach for making decisions in hypothesis tests
6. Compute power and type II error probability, and make sample size selection decisions for tests on means, variances, and proportions
7. Explain and use the relationship between confidence intervals and hypothesis tests

Motivating example

Suppose that an engineer is designing an air crew escape system that consists of an ejection seat and a rocket motor that powers the seat. The rocket motor contains a propellant, and for the ejection seat to function properly, the propellant should have a mean burning rate of 50 cm/sec.

If the burning rate is too low, the ejection seat may not function properly, leading to an unsafe ejection and possible injury of the pilot. Higher burning rates may imply instability in the propellant or an ejection seat that is too powerful, again leading to possible pilot injury.

So the practical engineering question that must be answered is: Does the mean burning rate of the propellant equal 50 cm/sec, or is it some other value (either higher or lower)?

This type of question can be answered using a statistical technique called **hypothesis testing**.

Hypothesis Testing

In the previous chapter, we illustrated how to construct a confidence interval estimate of a parameter from sample data. However, many problems in engineering require that we decide which of two competing claims or statements about some parameter is true. The statements are called **hypotheses**, and the decision-making procedure is called **hypothesis testing**.

Statistical Hypothesis

A **statistical hypothesis** is a statement about the parameters of one or more populations

Hypothesis Testing

Consider the air crew escape system. Suppose that we are interested in the burning rate of the solid propellant. Burning rate is a random variable that can be described by a probability distribution. Suppose that our interest focuses on the mean burning rate (a parameter of this distribution). Specifically, we are interested in deciding whether or not the mean burning rate is 50 centimeters per second. We may express this formally as

$$H_0 : \mu = 50 \text{ cm/sec versus } H_1 : \mu \neq 50 \text{ cm/sec.}$$

The statement H_0 is called the **null hypothesis**. This is a claim that is initially assumed to be true. The statement H_1 is called the **two-tailed (or two-sided) alternative hypothesis** which contradicts the null hypothesis.

Hypothesis Testing

In some situations, we may wish to formulate a **one-sided alternative hypothesis**, as in

$$H_0 : \mu = 50 \text{ cm/sec versus } H_1 : \mu < 50 \text{ cm/sec}$$
(left-tailed alternative hypothesis)

or $$H_0 : \mu = 50 \text{ cm/sec versus } H_1 : \mu > 50 \text{ cm/sec}$$
(right-tailed alternative hypothesis)

We will always state the null hypothesis as an equality claim in this course. However, when the alternative hypothesis is stated with the $<$ sign, the implicit claim in the null hypothesis can be taken as $\geq$ and when the alternative hypothesis is stated with the $>$ sign, the implicit claim in the null hypothesis can be taken as $\leq$.

Tests of Statistical Hypotheses

To illustrate the general concepts, consider the propellant burning rate problem introduced earlier:

$$H_0 : \mu = 50 \text{ cm/sec versus } H_1 : \mu \neq 50 \text{ cm/sec.}$$

Suppose that a sample of $n = 10$ specimens is tested and that the sample mean burning rate $\bar{x}_n$ is observed. The sample mean is an estimate of the true population mean μ.

- A value of the sample mean $\bar{x}_n$ that falls close to the hypothesized value of $\mu = 50$ cm/sec does not conflict with the null hypothesis that the true mean μ is really 50 cm/sec.
- On the other hand, a sample mean that is considerably different from 50 cm/sec is evidence in support of the alternative hypothesis H_1.

Thus, the sample mean is the test statistic in this case.

Tests of Statistical Hypotheses (Critical region, acceptance region, critical values

The sample mean can take on many different values. Suppose that

- if $48.5 \leq \bar{x}_n \leq 51.5$, we will not reject the null hypothesis $H_0 : \mu = 50$
- if either $\bar{x}_n < 48.5$ or $\bar{x}_n > 51.5$, we will reject H_0 in favor of the alternative hypothesis $H_1 : \mu \neq 50$.

The values of $\bar{x}_n$ that are less than 48.5 and greater than 51.5 constitute the **critical region** (or **rejection region**) for the test; all values that are outside the critical region form a region for which we will fail to reject H_0. By convention, this is usually called the **acceptance region**. The boundaries between the critical regions and the acceptance region are called the **critical values**.

In our example, the critical values are 48.5 and 51.5. It is customary to state conclusions relative to the null hypothesis H_0. Therefore, we reject H_0 in favor of H_1 if the test statistic falls in the critical region and fails to reject H_0 otherwise.

Type I, II errors

This decision procedure can lead to either of two wrong conclusions.

- The true mean burning rate **could** be equal to 50 cm/sec. However, for the randomly selected propellant specimens that are tested, we **could** observe a value of the test statistic $\bar{x}_n$ that falls into the critical region. We would then reject the null hypothesis H_0 in favor of the alternate H_1 when, in fact, H_0 is really true. This type of wrong conclusion is called a **type I error**.
- The true mean burning rate **could** be different from 50 cm/sec, yet the sample mean $\bar{x}_n$ **could** fall in the acceptance region. In this case, we would fail to reject H_0 when H_0 is false, and this leads to the other type of error, called a **type II error**.

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

Probability of Type I Error

Because our decision is based on random variables (e.g., $\bar{X}_n$), probabilities can be associated with the type I and type II errors. The probability of making a type I error is denoted by the Greek letter α.

Probability of Type I Error

$$\alpha = P(\text{Type I Error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}).$$

Back to the previous example, suppose the population is normal and the population standard deviation is $\sigma = 2.5$ cm/sec, then based on a sample of size $n = 10$,

$$\begin{aligned}\alpha &= P(\bar{X}_n < 48.5 \text{ or } \bar{X}_n > 51.5 \text{ when } \mu = 50) \\ &= normal.probability(-Inf, 48.5, 50, 2.5/\sqrt{10}) \\ &\quad + normal.probability(51.5, Inf, 50, 2.5/\sqrt{10}) = 0.0574.\end{aligned}$$

where by the normality and $\mu = 50$, $\bar{X}_n \sim N(\mu = 50, \sigma^2/n = 2.5^2/10)$.

Probability of Type II Error and Power

Probability of Type II Error (or known as β-error)

$$\beta = P(\text{Type II Error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}).$$

The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true; i.e., $1-\beta$.

Back to the previous example, suppose the population is normal and the population standard deviation is $\sigma = 2.5$ cm/sec, then based on a sample of size $n = 10$, the probability of Type II Error when $\mu = 52$ is

$$\begin{aligned}\beta &= P(48.5 \leq \bar{X}_n \leq 51.5 \text{ when } \mu = 52) \\ &= \mathit{normal.probability}(48.5, 51.5, 52, 2.5/\sqrt{10}) = 0.2643.\end{aligned}$$

where by the normality and $\mu = 52$, $\bar{X}_n \sim N(\mu = 52, \sigma^2/n = 2.5^2/10)$. The **power** of the test at $\mu = 52$ is $1 - 0.2643 = 0.7357$.

Summary

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	$\boxed{0.5000}$	$\boxed{0.9705}$
$48.81 < \bar{x} < 51.19$	16	$\boxed{0.0576}$	$\boxed{0.0966}$	$\boxed{0.8606}$
$48.42 < \bar{x} < 51.58$	16	$\boxed{0.0114}$	$\boxed{0.2515}$	$\boxed{0.9578}$

1. The size of the critical region, and consequently the probability of a type I error α, can always be reduced by appropriate selection of the critical values.
2. Type I and type II errors are related. A decrease in one always results in an increase in the other provided that the sample size n does not change.
3. An increase in n reduces β provided that α is held constant.
4. When H_0 is false, β increases as the true value of the parameter approaches the value hypothesized in the null hypothesis. The value of β decreases as the difference between the true mean and the hypothesized value increases.

Summary (continued)

Generally, we control/fix the type I error probability α to select the critical region. In this way, we can directly control the probability of wrongly rejecting H_0. We always think of rejection of the null hypothesis H_0 as a **strong conclusion**. We call the fixed α as the **significance level** of the test.

Below are the common steps (of a critical value approach)

1. Find a test statistics
2. Given the probability of type I error α, find a critical region.
3. If the test statistics falls in the critical region, reject H_0; otherwise, fail to reject.
4. Conclude the result.

Summary (continued)

We could also use a confidence-interval approach:

1. Based on the alternative hypothesis H_1 (two-tailed or left/right-tailed), compute a (two-sided or one-sided) confidence interval estimate of confidence level $100(1-\alpha)\%$.
2. If the confidence interval estimate covers the hypothesized value in H_0, we fail to reject H_0; otherwise, reject H_0.
3. Conclude the result.

Lastly, we could use a P-value approach. The P-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

1. Compute P-value based on the data.
2. If P-value less than α, reject H_0; otherwise, fail to reject H_0.
3. Conclude the result.

All these can be done using StatEngine!

Tests on the Mean of a Normal Distribution, Variance Known

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

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

Based on sample $x_1, \ldots, x_n$, the observed test statistic is $z_0 = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$

If the alternative hypothesis is $H_1 : \mu \neq \mu_0$. At significance level α, we reject H_0 if one of the following holds:

- $|z_0| > z_{\alpha/2}$ (critical-value approach),
- $\mu_0 \notin [\bar{x}_n \pm z_{\alpha/2}\sigma/\sqrt{n}]$ (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

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

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

Based on sample $x_1, \ldots, x_n$, the observed test statistic is $z_0 = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$

If the alternative hypothesis is $H_1 : \mu > \mu_0$. At significance level α, we reject H_0 if one of the following holds:

- $z_0 > z_\alpha$ (critical-value approach),
- $\mu_0 \notin [\bar{x}_n - z_\alpha \sigma/\sqrt{n}, +\infty)$ (confidence-interval approach) or $\mu_0 < \bar{x}_n - z_\alpha \sigma/\sqrt{n}$, the $100(1-\alpha)\%$ lower bound on μ (μ_0 is even smaller than the smallest confident guess of μ).
- the P-value $= P(Z > z_0) < \alpha$, where $Z \sim N(0,1)$ (P-value approach).

Tests on the Mean of a Normal Distribution, Variance Known

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

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

Based on sample $x_1, \dots, x_n$, the observed test statistic is $z_0 = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$

If the alternative hypothesis is $H_1 : \mu < \mu_0$. At significance level α, we reject H_0 if one of the following holds:

- $z_0 < -z_\alpha$ (critical-value approach),
- $\mu_0 \notin (-\infty, \bar{x}_n + z_\alpha \sigma/\sqrt{n}]$ (confidence-interval approach) or $\mu_0 > \bar{x}_n + z_\alpha \sigma/\sqrt{n}$, the $100(1-\alpha)\%$ upper bound on μ (μ_0 is even larger than the largest confident guess of μ),
- the P-value $= P(Z \leq z_0) < \alpha$, where $Z \sim N(0, 1)$ (P-value approach).

Probability of a Type II Error for Tests on the Mean, Variance Known

Type II error occurs when H_1 is true. Suppose the true mean value is $\mu = \mu_1$. Let $\delta = \mu_1 - \mu_0$.
If the alternative hypothesis is $H_1 : \mu \neq \mu_0$. The probability of the type II error of the previous two-tailed test is

$$\beta = P\left(-z_{\alpha/2} \leq \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2}, \text{ under } H_1 : \mu = \mu_1\right)$$

$$\text{Under } H1 : \bar{X}_n \sim N(\mu = \mu_1, \sigma^2/n)$$

$$= P\left(-z_{\alpha/2} \leq \frac{\bar{X}_n - \mu_1 + \mu_1 - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right)$$

$$= P\left(-z_{\alpha/2} \leq Z + \frac{\delta}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right)$$

$$= P\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma} \leq Z \leq z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right).$$

Probability of a Type II Error for Tests on the Mean, Variance Known

If the alternative hypothesis is $H_1 : \mu > \mu_0$. The probability of the type II error of the previous right-tailed test is

$$\beta = P\left(Z \leq z_\alpha - \frac{\delta\sqrt{n}}{\sigma}\right).$$

If the alternative hypothesis is $H_1 : \mu < \mu_0$. The probability of the type II error of the previous left-tailed test is

$$\beta = P\left(-z_\alpha - \frac{\delta\sqrt{n}}{\sigma} \leq Z\right).$$

Choice of Sample Size for Tests on the Mean, Variance Known

What is an appropriate sample size to control the probability of type II error β (or the power $1-\beta$) for given δ and α?

If the alternative hypothesis is $H_1 : \mu \neq \mu_0$. The sample size should be at least the smallest positive integer n such that

$$\beta \geq P\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma} \leq Z \leq z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right).$$

If the alternative hypothesis is $H_1 : \mu > \mu_0$. The sample size should be at least the smallest positive integer n such that

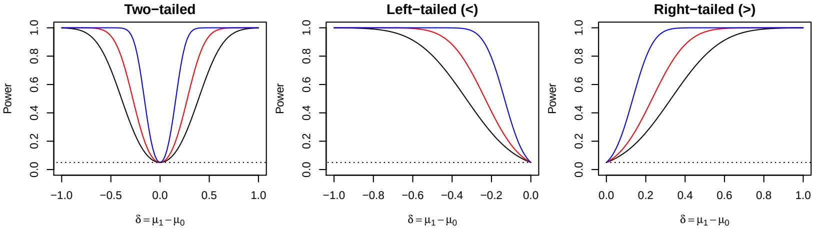
$$\beta \geq P\left(Z \leq z_{\alpha} - \frac{\delta\sqrt{n}}{\sigma}\right).$$

If the alternative hypothesis is $H_1 : \mu < \mu_0$. The sample size should be at least the smallest positive integer n such that

$$\beta \geq P\left(-z_{\alpha} - \frac{\delta\sqrt{n}}{\sigma} \leq Z\right).$$

We can find these numerically (StatEngine).

Summary of Tests on the Mean, Variance Known



Herein, $\alpha = 0.05$, $\sigma = 1$. Power curves: black $n = 25$, red $n = 50$, blue $n = 200$.

summary of the patterns

- Probability of Type I Error is always α, user-chosen.
- When the sample size n is fixed, the power of test ($1 - \beta$, where β is the Type II Error) increases as the difference (δ) between the true parameter value and the hypothesized value increases in favor of H_1. Easier to detect a stronger signal!
- When a non-zero difference ($\delta \neq 0$) between the true parameter value and the hypothesized value in favor of H_1 is fixed, the power of test ($1 - \beta$) increases as n increases. Easier to detect a signal with more samples!
- If $\delta = 0$, the hypothesized value is the true parameter value, the power of the test is always α no matter how large n is. Probability of Type I Error does not change with n.

Summary of Tests on the Mean, Variance Known

Z-test

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

Three types of H_1 ($\mu \neq, <, > \mu_0$)

Observed test statistic: $z_0 = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$

- Critical-value approach
- Confidence-interval approach
- P-value approach
- Power calculation
- Choice of sample size

StatEngine Summary of Tests on the Mean, Variance Known

Conduct the test:

```
#If data are available
Ztest(mu0=?,H1=?,alpha=?,sigma=?,sample=?)

#If statistics are provided
Ztest(mu0=?,H1=?,alpha=?,sigma=?,n=?,barx=?)
```

Calculate power $(1-\beta)$ of the test for given δ and n:

```
Ztest.power(H1=?,alpha=?,sigma=?,n=?,delta=?)
```

Calculate the minimum sample size to control β for given δ and α:

```
sample.size.Ztest(H1=?,sigma=?,alpha=?,beta=?,delta=?)
```

Example

A manufacturer produces crankshafts for an automobile engine. The crankshafts wear after 100,000 miles (0.0001 inch) is of interest because it is likely to have an impact on warranty claims. A random sample of $n = 15$ shafts is tested and $\bar{x}_n = 2.78$. It is known that $\sigma = 0.9$ and that wear is normally distributed.

(a) Test $H_0 : \mu = 3$ versus $H_1 : \mu \neq 3$ using $\alpha = 0.05$.
(b) What is the power of this test if $\mu = 3.25$?
(c) What sample size would be required to detect a true mean of 3.75 if we wanted the power to be at least 0.9?
(d) Explain how the question in part (a) could be answered by constructing a confidence interval on the μ.

Solution: 1. It is about μ; 2. Normal distribution; 3. $\sigma = 0.9$ is known. Bingo: Z-test!

(a) No data but statistics are provided.

```
Ztest(mu0=3,H1="two",alpha=0.05,sigma=0.9,n=15,barx=2.78)
```

Example (continued)

(a) No data but statistics are provided.

```
Ztest(mu0=3,H1="two",alpha=0.05,sigma=0.9,n=15,barx=2.78)
```

It also answers part (d).

```
The sample mean is 2.78 and sample size is 15

H1 is two-tailed: mu does not equal to mu0=3 . The results are:

1. Test statistic z0 is -0.9467293 , z_(alpha/2) is 1.959964,
Because -z_(alpha/2)<=z0<=z_(alpha/2), we fail to reject H0
at significance level 0.05

2. A 95% two-tailed confidence interval for the population mean is
[2.324546 , 3.235454] which contains the hypothesized value mu0=3,
we fail to reject H0 at significance level 0.05

3. The P-value is 0.3437768 which is not smaller than alpha= 0.05,
so we fail to reject H0 at significance level 0.05
```

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

Example (continued)

(b) Power calculation for $n = 15$ and $\delta = 3.25 - 3 = 0.25$:

```
Ztest.power(H1="two",alpha=0.05,sigma=0.9,n=15,delta=3.25-3)

H1 is two-tailed

The probability of the Type II error of this test at delta=mu1-mu0= 0.25
is 0.8104889 and the associated power is 0.1895111
```

(c) Sample size calculation for $\beta = 1 - 0.9 = 0.1$, $\delta = 3.75 - 3$, and $\alpha = 0.05$.

```
sample.size.Ztest(H1="two",sigma=0.9,alpha=0.05,beta=1-0.9,delta=3.75-3)

At significance level alpha= 0.05 , we need at least n= 16 to
make the power of this test at delta= 0.75 be at least 0.9
```

(d) See the solution to part (a).

Tests on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

Test statistic: $T_0 = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$

Based on sample $x_1, \ldots, x_n$, the observed test statistic is $t_0 = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$

If the alternative hypothesis is $H_1 : \mu \neq \mu_0$. At significance level α, we reject H_0 if one of the following holds:

- $|t_0| > t_{n-1,\alpha/2}$ (critical-value approach),
- $\mu_0 \notin [\bar{x}_n \pm t_{n-1,\alpha/2} s_n/\sqrt{n}]$ (confidence-interval approach),
- the P-value $= 2[1 - P(T_{n-1} \leq |t_0|)] < \alpha$, where $T_{n-1} \sim t(n-1)$ (P-value approach).

Tests on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

Test statistic: $T_0 = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$

Based on sample $x_1, \ldots, x_n$, the observed test statistic is $t_0 = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$

If the alternative hypothesis is $H_1 : \mu > \mu_0$. At significance level α, we reject H_0 if one of the following holds:

- $t_0 > t_{n-1,\alpha}$ (critical-value approach),
- $\mu_0 \notin [\bar{x}_n - t_{n-1,\alpha} s_n/\sqrt{n}, +\infty)$ (confidence-interval approach) or $\mu_0 < \bar{x}_n - t_{n-1,\alpha} s_n/\sqrt{n}$, the $100(1-\alpha)\%$ lower bound on μ,
- the P-value $= P(T_{n-1} > t_0) < \alpha$ (P-value approach).

Tests on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

Test statistic: $T_0 = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$

Based on sample $x_1, \ldots, x_n$, the observed test statistic is $t_0 = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$

If the alternative hypothesis is $H_1 : \mu < \mu_0$. At significance level α, we reject H_0 if one of the following holds:

- $t_0 < -t_{n-1,\alpha}$ (critical-value approach),
- $\mu_0 \notin (-\infty, \bar{x}_n + t_{n-1,\alpha} s_n/\sqrt{n}]$ (confidence-interval approach) or $\mu_0 > \bar{x}_n + t_{n-1,\alpha} s_n/\sqrt{n}$, the $100(1-\alpha)\%$ upper bound on μ,
- the P-value $= P(T_{n-1} \leq t_0) < \alpha$ (P-value approach).

Power and Choice of Sample Size for Tests on the Mean of a Normal Distribution, Variance Unknown

T-test

Null hypothesis: $H_0 : \mu = \mu_0$ for a hypothesized value μ_0.

Three types of H_1 ($\mu \neq, <, > \mu_0$)

- Power calculation
- Choice of sample size

To exactly calculate the power and the minimum sample size, we must know σ which is unknown in the use of a T-test. Therefore, we use the sample standard deviation s_n as an estimate to approximate the results.

In addition, like T-interval, T-test is also robust to the normality assumption. **In other words, if the distribution is not normal, one can still use a T-test**. But the result are all approximation; e.g., the probability of Type I error is controlled to be approximately α.

StatEngine Summary of Tests on the Mean, Variance Unknown

Conduct the test:

```
#If data are available
Ttest(mu0=?,H1=?,alpha=?,sample=?)

#If statistics are provided
Ttest(mu0=?,H1=?,alpha=?,n=?,barx=?,s=?)
```

Calculate power $(1-\beta)$ of the test for given δ and n:

```
sn=sd(data)
Ttest.power(H1=?,est.sigma=sn,alpha=?,n=?,delta=?)
```

Calculate the minimum sample size to control β for given δ and α:

```
sample.size.Ttest(H1=?,est.sigma=sn, beta=?,delta=?,alpha=?)
```

Example

The sodium content of twenty 300-gram boxes of organic cornflakes was determined. The data (in milligrams) are as follows:
x=c(131.15, 130.69, 130.91, 129.54, 129.64, 128.77, 130.72, 128.33, 128.24, 129.65, 130.14, 129.29, 128.71, 129.00, 129.39, 130.42, 129.53, 130.12, 129.78, 130.92)

(a) Can you support a claim that mean sodium content of this brand of cornflakes differs from 130 milligrams? Use $\alpha = 0.05$. Find the P-value.

(b) Check that sodium content is normally distributed.

(c) Compute the power of the test if the true mean sodium content is 130.5 milligrams.

(d) What sample size would be required to detect a true mean sodium content of 130.1 milligrams if you wanted the power of the test to be at least 0.75?

(e) Explain how the question in part (a) could be answered by constructing a two-sided confidence interval on the mean sodium content.

Example (solution)

(a) Consider $H_0 : \mu = 130$ versus $H_1 : \mu \neq 130$. We do not know the variance of the population. Known T-test is robust to distributional assumption. We will use T-test:

```
Ttest(mu0=130,H1="two",alpha=0.05,sample=x)
```

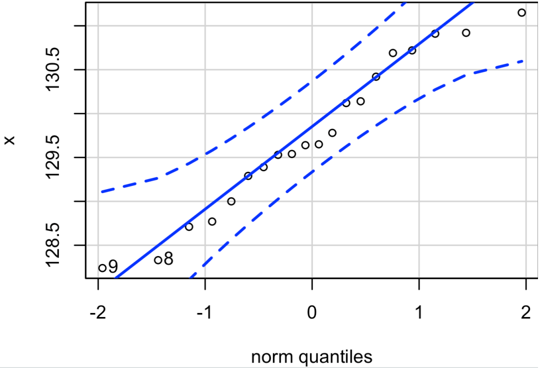
```
The sample mean is 129.747 sample standard deviation is 0.8764288 , and sample

H1 is two-sided: mu does not equal to mu0= 130 . The results are:

1. Test statistic t0 is -1.290978 , t_(n-1,alpha/2) is 2.093024,
Because |t0|<=t_(n-1,alpha/2), we fail to reject H0 at significance level 0.05

2. A 95% two-sided confidence interval for the population mean is
[ 129.3368 , 130.1572 ] which contains the hypothesized value mu0= 130 ,
so we fail to reject H0 at significance level 0.05

3. The P-value is 0.2121988 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 (solution continued)

(b) We use QQ plot to check normality

```
data.summary(x)
```



Most dots are fitted well by the line. Thus, we could conclude that the data are from a normal distribution, which further supports our use of the T-test in (a).

Example (solution continued)

(c) Find the power when $\mu_1 = 130.5$. From (a), we know $\mu_0 = 130$ yielding $\delta = \mu_1 - \mu_0 = 0.5$

```
# The power calculation requires an estimate of
# the population standard deviation, we use
# the sample standard deviation which can be obtained
# by sd(x). The sample size n=length(x)

Ttest.power(H1="two",est.sigma=sd(x),alpha=0.05,
    n=length(x),delta=0.5)

H1 is two-sided

The probability of the Type II error of this test at
delta=mu1-mu0= 0.5 is 0.3224236
and the associated power is 0.6775764
```

Example (solution continued)

(d) Now we have $\mu_1 = 130.1$ and $\beta = 1 - 0.75 = 0.25$. From (a), we know $\mu_0 = 130$ yielding $\delta = \mu_1 - \mu_0 = 0.1$

```
# The sample size calculation requires an estimate of
# the population standard deviation, we use
# the sample standard deviation
# which can be obtained by sd(x).

sample.size.Ttest(H1="two",est.sigma =sd(x),
    beta=0.25,delta=0.1,alpha=0.05)

At significance level alpha= 0.05 , we need at least
n= 536 to achieve a power >= 0.75 of this test
at delta= 0.1
When n= 535 , the power is 0.7499781
When n= 536 , the power is 0.7507622
When n= 537 , the power is 0.7515444
```

(e) See the StatEigine output in part (a).

Large-sample tests on the Mean of a Distribution

Similarly to the large-sample Z-interval, we can use the central limit theorem to develop a large-sample Z-test for $H_0 : \mu = \mu_0$ versus the three types of alternative hypotheses.

- Advantages are to lift the normality assumption and to allow σ to be unknown.
- A disadvantage is the need of a large sample size (e.g., $n \geq 25$) and everything is approximated.

However, we know T-test is robust to the normality assumption and is designed for unknown σ. Thus, we can always use T-test whenever large-sample Z-test can be used. One must acknowledge that T-test is always more conservative (less powerful) than large-sample Z-test; i.e., T-test often requires a larger sample size to reach the same power when compared to large-sample Z-test; or when the sample size is the same, the power of T-test is smaller than the one of large-sample Z-test.

Tests on the Variance and Standard Deviation of a Normal Distribution

- An automated filling machine is used to fill bottles with liquid detergent. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled.
- If the standard deviation of hole diameter exceeds 0.01 millimeters, there is an unacceptably high probability that the rivet will not fit.

In these applications, we want to make inference about the population variance σ^2/population standard deviation σ. In particular, we now focus on these hypotheses:

$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 \neq \sigma_0^2 \text{ (two-tailed)}$$
$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 > \sigma_0^2 \text{ (right-tailed)}$$
$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 < \sigma_0^2 \text{ (left-tailed)}$$

Tests on the Variance and Standard Deviation of a Normal Distribution

Basic reasoning:
We know the sample variance S_n^2 is a good estimator of the true population variance σ^2. Thus, if based on the observed sample, s_n^2 is close to σ_0^2 or $\frac{s_n^2}{\sigma^2}$ is close to 1, then the data suggests the hypothesized value σ_0^2 is close to the true value σ^2, and thus we do not reject H_0. This reasoning leads to our consideration of test statistic

$$\boldsymbol{X}_0^2 = \frac{(n-1)S_n^2}{\sigma_0^2}$$

When H_0 is true (control the type I error), $\boldsymbol{X}_0^2$ follows the chi-square distribution with $n-1$ degrees of freedom. Thus:

Tests on the Variance and Standard Deviation of a Normal Distribution

Chi-square test (need normality):

Null hypothesis: $H_0 : \sigma^2 = \sigma_0^2$ for a hypothesized value σ_0^2.

Test statistic: $\boldsymbol{X}_0 = \frac{(n-1)S_n^2}{\sigma_0^2}$

Based on sample $x_1, \dots, x_n$, the observed test statistic is

$$\boldsymbol{x}_0^2 = \frac{(n-1)s_n^2}{\sigma_0^2}$$

- If $H_1 : \sigma^2 \neq \sigma_0^2$. At significance level α, we reject H_0 if $\boldsymbol{x}_0^2 < \chi^2_{n-1,1-\alpha/2}$ or $\boldsymbol{x}_0^2 > \chi^2_{n-1,\alpha/2}$;
- if $H_1 : \sigma^2 < \sigma_0^2$. At significance level α, we reject H_0 if $\boldsymbol{x}_0^2 < \chi^2_{n-1,1-\alpha}$;
- if $H_1 : \sigma^2 > \sigma_0^2$. At significance level α, we reject H_0 if $\boldsymbol{x}_0^2 > \chi^2_{n-1,\alpha}$.

Tests on the Variance and Standard Deviation of a Normal Distribution

Chi-square test:

- Besides the above critical value approach,
- we also have a confidence-interval approach,
- and the P-value approach.
- Furthermore, we can compute the power
- and determine the minimum required sample size.

A power is 1 minus the probability of a Type II error. A Type II error occurs when H_1 is true. Suppose the true population variance is $\sigma_1^2 \neq \sigma_0^2$, we denote

$$\lambda = \frac{\sigma_1}{\sigma_0},$$

and will use this λ to compute the power and determine the minimum required sample size.

StatEngine Summary of Tests on the Variance and Standard Deviation of a Normal Distribution

Conduct the test:

```
#If data are available
Chi2test(sigma0=?,H1=?,alpha=?,sample=?)

#If statistics are provided
Chi2test(sigma0=?,H1=?,alpha=?,n=?,s=?)
```

Calculate power $(1-\beta)$ of the test for given λ and *n*:

```
Chi2test.power(H1=?,alpha=?,n=?,lambda=?)
```

Calculate the minimum sample size to control β for given λ and α:

```
sample.size.Chi2test(H1=?,beta=?,lambda=?,alpha=?)
```

Remark: Normality is important. If normality does not hold, using these might give you wrong inference. If possible, use QQ-plot to check normality (data.summary($\cdot$) in StatEngine).

Example

An automated filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s_n^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

Solution: It is about the population variance σ^2. We use Chi.square test, which requires normality. Fortunately, normality is assumed by the question. No data are provided, we do have $s_n^2 = 0.0153$ where $n = 20$. The hypotheses are

$$H_0 : \sigma^2 = 0.01 \text{ versus } H_1 : \sigma^2 > 0.01$$

where $\sigma_0^2 = 0.01$, and the alternative is right-tailed. We set $\alpha = 0.05$.

Example (continued)

Using StatEngin

```
Chi2test(sigma0=sqrt(0.01),H1="right",alpha=0.05,n=20,s=sqrt(0.0153))

df= 19 , sample variance is 0.0153 , sample standard deviation
is 0.1236932 , and sample size is 20

H1 is right-tailed: sigma^2 is more than sigma0^2= 0.01 .
The results are:

1. Test statistic x02 is 29.07 , Chi2_(n-1,alpha) is 30.14353 ,
Because x02<= Chi2_(n-1,alpha),
we fail to reject H0 at significance level 0.05

2. A 95% one-sided confidence interval for the population
variance is [ 0.009643861 , Inf )
which contains the hypothesized value sigma0^2= 0.01 ,
so we fail to reject H0 at significance level 0.05

3. The P-value is 0.064892 which is not smaller than alpha= 0.05 ,
so we fail to reject H0 at significance level 0.05
```

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

Example (continued)

Suppose that if the true standard deviation of the filling process exceeds the hypothesized value σ_0 by 25%, what is the power of the previous test?

Solution: This asks for a power calculation of the chi.square test we had conducted. In the previous test, we have a sample of size $n = 20$. Now the true parameter is $\sigma_1 = (1 + 0.25)\sigma$, what is the power? We use StatEngine, which needs the type of H_1 (right-tailed), the significance level $\alpha = 0.05$, the sample size $n = 20$, and the value of $\lambda = \sigma_1/\sigma_0 = 1.25$.

```
Chi2test.power(H1="right",alpha=0.05,n=20,lambda=1.25)

H1 is right-tailed (>)

The probability of the Type II error of this test at
lambda=sigma1/sigma0= 1.25 is 0.5617379
and the associated power is 0.4382621
```

The power of the right-tailed test we have conducted when $n = 20$ is 0.4383.

Example (continued)

Suppose that if the true standard deviation of the filling process exceeds the hypothesized value σ_0 by 25%, we would like to detect this with probability at least 0.8. Is the sample size of $n = 20$ adequate?

Solution: This asks for a sample size calculation. It says if $\sigma_1 = (1+0.25)\sigma_0$, we want to control the Type II error β by $1-0.8 = 0.2$. Is $n = 20$ adequate? We use StatEngine to find the minimum sample size. The StatEngine needs the type of H_1 (right-tailed), the required value of $\beta = 0.2$, the value of $\lambda = \sigma_1/\sigma_0 = 1.25$, and the significance level $\alpha = 0.05$.

```
sample.size.Chi2test(H1="right",beta=0.2,lambda=1.25,alpha=0.05)
```

```
At significance level alpha= 0.05 , we need at least n= 61 to
achieve a power >= 0.8 of this test
at lambda=sigma1/sigma0= 1.25
When n= 60 , the power is 0.7954305
When n= 61 , the power is 0.8008053
When n= 62 , the power is 0.8060504
```

We need at least 61 samples. Thus $n = 20$ is not adequate!

Example 2

If the standard deviation of hole diameter exceeds 0.01 millimeters, there is an unacceptably high probability that the rivet will not fit. Suppose that $n = 15$ and $s_n = 0.008$ millimeter.

(a) Is there strong evidence to indicate that the standard deviation of hole diameter exceeds 0.01 millimeter? Use $\alpha = 0.01$. State any necessary assumptions about the underlying distribution of the data. Find the P-value for this test.

(b) Suppose that the actual standard deviation of hole diameter exceeds the hypothesized value by 50%. What is the probability that this difference will be detected by the test described in part (a)?

(c) If σ is really as large as 0.0125 millimeters, what sample size will be required to detect this with power of at least 0.8?

Example 2 (Solution)

(a) Right-tailed alternative with $\sigma_0 = 0.01^2$

```
Chi2test(sigma0=0.01,H1="right",alpha=0.01,n=15,s=0.008)
```

We need to assume the samples are from a normal distribution.
Conclusion:...

(b) You should be able to identify $\lambda = 1.5$

```
Chi2test.power(H1="right",alpha=0.01,n=15,lambda=1.5)
```

(c) Now $\lambda = 0.0125/0.01 = 1.25$, $\beta = 1 - 0.8 = 0.2$.

```
sample.size.Chi2test(H1="right",beta=0.2,lambda=1.25,alpha=0.01)
```

Try these yourself!

Tests on a Population Proportion

It is often necessary to test hypotheses on a population proportion:

$$H_0: p = p_0 \text{ versus } H_1: p \neq p_0 \text{ (two-tailed)}$$
$$H_0: p = p_0 \text{ versus } H_1: p < p_0 \text{ (left-tailed)}$$
$$H_0: p = p_0 \text{ versus } H_1: p > p_0 \text{ (right-tailed)}$$

suppose that a random sample of size n has been taken from a large (possibly infinite) population and that $X(\leq n)$ observations in this sample belong to a class of interest.

We know the sample proportion $\widehat{p} = \frac{X}{n}$ is a good estimator of the true population proportion p. Thus, if based on the observed sample, $\widehat{p}$ is close to p_0, then the data suggests the hypothesized value p_0 is close to the true value p, and thus we do not reject H_0. This reasoning leads to our consideration of test statistic

$$Z_0 = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Tests on a Population Proportion

One-Sample Proportion Z-test (a large-sample test):
Require: $n\widehat{p} \geq 5$ and $n(1-\widehat{p}) \geq 5$.

Null hypothesis: $H_0 : p = p_0$ for a hypothesized value p_0.

The observed test statistic: $z_0 = \frac{\widehat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

- If $H_1 : p \neq p_0$. At significance level α, we reject H_0 if $|z_0| > z_{\alpha/2}$;
- if $H_1 : p < p_0$. At significance level α, we reject H_0 if $z_0 < -z_\alpha$;
- if $H_1 : p > p_0$. At significance level α, we reject H_0 if $z_0 > z_\alpha$.

One-Sample Proportion Z-test:

- Besides the above critical value approach,
- we also have a confidence-interval approach,
- and the P-value approach.
- Furthermore, we can **approximate** the power
- and **approximate** the minimum required sample size.

In this test, the confidence-interval approach **might** give a different conclusion than the other two approaches. But this difference is very minor, especially when sample size n is large.

When approximating the power, which is 1 minus the probability of a Type II error and a Type II error occurs when H_1 is true, we suppose the true population proportion is p_1. The p_0, p_1, and n are needed when computing a power and minimum required sample size for this test.

StatEngine Summary of Tests on the Variance and Standard Deviation of a Normal Distribution

Conduct the test:

```
Proptest(p0=?,H1=?,alpha=?,n=?,X=?)
```

Approximate power $(1 - \beta)$ of the test for given δ and n:

```
Proptest.power(H1=?,alpha=?,p0=?,p1=?,n=?)
```

Approximate the minimum sample size to control β for given p_1, p_0, and α:

```
sample.size.Proptest(H1=?,beta=?,alpha=?,p0=?,p1=?)
```

Example

A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 300 devices and finds that 7 of them are defective. Can the manufacturer demonstrate process capability for the customer?

Solution: p is the population defective rate. We want to test $H_0 : p \geq 0.05$ versus $H_1 : p < 0.05$. Thus $p_0 = 0.05$. We also know $n = 300$, $X = 7$ and $\alpha = 0.05$. Then $\widehat{p} = 7/300$ and we can verify that $n\widehat{p} \geq 5$ and $n(1 - \widehat{p}) \geq 5$. Now we can perform the left-tailed One-Sample Proportion Z-test:

```
Proptest(p0=0.05, H1="left",alpha=0.05,n=300,X=7)
```

Example (continued)

Perform the left-tailed One-Sample Proportion Z-test:

```
Proptest(p0=0.05, H1="left",alpha=0.05,n=300,X=7)

The sample proportion is 0.02333333 and sample size is 300

H1 is left-tailed: p is less than p0= 0.05 . The results are:

1. Test statistic z0 is -2.119252 , 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
( -Inf , 0.03766934 ] which does not contain the hypothesized
value p0= 0.05 , we reject H0 at significance level 0.05

3. The P-value is 0.0170346 which is smaller than alpha= 0.05 ,
so we reject H0 at significance level 0.05
```

Conclusion: at significance level $\alpha = 0.05$, the data provide sufficient evidence to reject H_0; i.e., the process is capable.

Example (continued)

Suppose that its process fallout rate is really $p = 0.03$. What is the β-error for a test of process capability that uses $n = 300$ and $\alpha = 0.05$?

Solution: This is a Typer II Error (1-Power) calculation, where $p_1 = 0.03$, $p_0 = 0.05$, $n = 300$, left-tailed alternative at $\alpha = 0.05$.

```
Proptest.power(H1="left",alpha=0.05,p0=0.05,p1=0.03,n=300)
```

```
H1 is left-tailed

The probability of the Type II error of this test at p1-p0= -0.02
is 0.5282211 and the associated power is 0.4717789
```

Example (continued)

Suppose that the semiconductor manufacturer was willing to accept a β-error as large as 0.10 if the true value of the process fraction defective was $p = 0.03$. If the manufacturer continues to use $\alpha = 0.05$, what sample size would be required?

Solution: This is a sample size calculation, where $p_1 = 0.03$, $p_0 = 0.05$, $\beta = 0.1$, left-tailed alternative at $\alpha = 0.05$.

```
sample.size.Proptest(H1="left",beta=0.1,alpha=0.05,p0=0.05,p1=0.03)
```

```
At significance level alpha= 0.05 , we need at least n= 833
to achieve a power >= 0.9 of this test at p1-p0= -0.02
When n= 832 , the power is 0.899778
When n= 833 , the power is 0.9001346
When n= 834 , the power is 0.9004902
```

StatEngine Summary of One-Sample Tests

Z-tests on μ (Normality, known σ)

```
Ztest(mu0=?,H1=?,alpha=?,sigma=?,sample=?) #If data are available
Ztest(mu0=?,H1=?,alpha=?,sigma=?,n=?,barx=?) #If statistics are provided
Ztest.power(H1=?,alpha=?,sigma=?,n=?,delta=?)
sample.size.Ztest(H1=?,sigma=?,alpha=?,beta=?,delta=?)
```

T-tests on μ (unknown σ)

```
Ttest(mu0=?,H1=?,alpha=?,sample=?) #If data are available
Ttest(mu0=?,H1=?,alpha=?,n=?,barx=?,s=?) #If statistics are provided
sn=sd(data)
Ttest.power(H1=?,est.sigma=sn,alpha=?,n=?,delta=?)
sample.size.Ttest(H1=?,est.sigma=sn, beta=?,delta=?,alpha=?)
```

StatEngine Summary of One-Sample Tests

Chi-square tests on σ^2 or σ (Normality)

```
Chi2test(sigma0=?,H1=?,alpha=?,sample=?) #If data are available
Chi2test(sigma0=?,H1=?,alpha=?,n=?,s=?) #If statistics are provided
Chi2test.power(H1=?,alpha=?,n=?,lambda=?)
sample.size.Chi2test(H1=?,beta=?,lambda=?,alpha=?)
```

One-Sample Proportion Z-tests on p ($n\widehat{p} \geq 5$, $n(1-\widehat{p}) \geq 5$)

```
Proptest(p0=?,H1=?,alpha=?,n=?,X=?)
Proptest.power(H1=?,alpha=?,p0=?,p1=?,n=?)
sample.size.Proptest(H1=?,beta=?,alpha=?,p0=?,p1=?)
```

A Summary of Pattern of Hypothesis Testing

- Probability of Type I Error is always α, user-chosen.
- When the sample size n is fixed, the power of test increases as the difference between the true parameter value and the hypothesized value increases in favor of H_1. Easier to detect a stronger signal!
- When a non-zero difference between the true parameter value and the hypothesized value in favor of H_1 is fixed, the power of test increases as n increases. Easier to detect a signal with more samples!
- If the difference is zero or the hypothesized value is the true parameter value, the power of the test reduces to the Probability of Type I Error which is always α no matter how large n is. Probability of Type I Error does not change with n.