1 Introduction

1.1 Motivation

You often come across data in various sizes and formats. It is important that you can use some *statistical models* to analyze these data with the aid of a *statistical software* in order to gain useful information out of the data.

Data are the raw material for statisticians

Some motivating examples:

Example: (Flu vaccine) A flu vaccine is known to be 20% effective in the second year after inoculation. To determine if a new vaccine is more effective 12 people are chosen at random and inoculated. Five of those (42%) receiving the new vaccine do not contact the virus in the second year after vaccination.

A researcher may ask:

Is the new vaccine superior to the old one?

Example: (Beer contents) A brand of beer claims its beer content is 375 (in millilitres) on the label. A sample of 40 bottles of the beer gave a sample average of 373.9 and a standard deviation of 2.5.

A consumer may ask:

Is there evidence that the mean content of the beer is less than 375 mL as claimed on the label?

Example: (Height comparison) A fourth grade class has 10 girls and 13 boys. The children's heights are recorded on their 10th birthday as follows:

Boys: 135.3 137.0 136.0 139.7 136.5 139.2 138.8 139.6 140.0 142.7 135.5 134.9 139.5;

Girls: 140.3 134.8 138.6 135.1 140.0 136.2 138.7 135.5 134.9 140.0

An interesting question is:

Is there evidence that girls are taller than boys on their 10th birthday?

Example: (Body temperature) The following dataset contains measurements of body temperature (in degree Celsius) over a certain time period.

```
> beav1$temp
  [1] 36.33 36.34 36.35 36.42 36.55 36.69 36.71 36.75 36.81 36.88 36.89 36.91
 [13] 36.85 36.89 36.89 36.67 36.50 36.74 36.77 36.76 36.78 36.82 36.89 36.99
 [25] 36.92 36.99 36.89 36.94 36.92 36.97 36.91 36.79 36.77 36.69 36.62 36.54
 [37] 36.55 36.67 36.69 36.62 36.64 36.59 36.65 36.75 36.80 36.81 36.87 36.87
 [49] 36.89 36.94 36.98 36.95 37.00 37.07 37.05 37.00 36.95 37.00 36.94 36.88
 [61] 36.93 36.98 36.97 36.85 36.92 36.99 37.01 37.10 37.09 37.02 36.96 36.84
 [73] 36.87 36.85 36.85 36.87 36.89 36.86 36.91 37.53 37.23 37.20 37.25 37.20
 [85] 37.21 37.24 37.10 37.20 37.18 36.93 36.83 36.93 36.83 36.80 36.75 36.71
 [97] 36.73 36.75 36.72 36.76 36.70 36.82 36.88 36.94 36.79 36.78 36.80 36.82
[109] 36.84 36.86 36.88 36.93 36.97 37.15
```

We may ask:

Is there evidence that the observations do not follow a normal population?

1.2 Statistical tests

This course attempts to answer these typical yes/no questions through the following steps:

1. **Hypotheses:** $H_0: \ \theta = \theta_0 \quad \text{vs} \quad H_1: \ \theta > \theta_0$, ...
2. **Test statistic:** $T = f(X_1, X_2, ..., X_n)$
3. **Assumptions:** Sample $X_1, X_2, ..., X_n \sim F_\theta$
4. ***P*-value:** p-value $= \Pr(T \geq t_0)$
5. **Decision:** If p-value < 0.05, there is evidence against H_0.
   If p-value > 0.05, the data are consistent with H_0.

1.3 Review

Before discussing statistical tests, let's review some basic knowledge.

1.3.1 Definitions

1. Statistics concerns itself mainly with conclusions and predictions resulting from *chance outcomes* that occur in carefully planned experiments or investigations. E.g. the flu vaccine inoculation or the beer content measurements.

2. A *population* is a collection of all possible measurements or observations of interest, e.g. all people targeted for the flu vaccine.

3. A *census* is complete evaluation of all members of the population and it provides all of the desired information.

4. The distribution of all population measurements is called a *population distribution*.

5. The quantities that determine the exact shape of the population distribution are called *parameters*, denoted by $\boldsymbol{\theta}$. Examples of parameters include the *mean* μ and *variance* σ^2.

6. A *census* is required to obtain the true values of parameters but such an extensive evaluation is often infeasible due to cost, time, manpower etc.

7. Instead statisticians use a representative *sample* from the population to infer the true parameters of the population. The sample is a *subset* of the population. E.g. 20 inoculated people or 40 bottles of beer.

8. The selection of a sample must be *random* so that each sample has *known* or even *same* probability of being selected. A random sample of size n, denoted by $X_1, X_2, \cdots, X_n$, is a sequence of *independent*

and *identically distributed* (iid) *random variable* (rv) with the same population distribution.

1.3.2 Some popular distributions

1. **The binomial distribution (P.103-110)**

   If a *discrete* rv X which counts the number of successes out of n independent trials each with a success probability p follows a binomial distribution with parameters n and p, denoted by $X \sim \mathcal{B}(n, p)$, its *probability mass function* (pmf), mean and variance are respectively:

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x (1-p)^{n-x}, \ x = 0, 1, \cdots, n, \\ E(X) &= np \ \text{ and } \ Var(X) = np(1-p) \end{aligned}$$

2. **The Poisson distribution (P.222-235)**

   If a *discrete* rv X which counts the number of certain events follows a Poisson distribution with parameter λ, denoted by $X \sim \mathcal{P}oi(\lambda)$, its pmf, mean and variance are respectively:

$$\begin{aligned} P(X = x) &= \frac{e^{-\lambda}\lambda^x}{x!}, \ x = 0, 1, \cdots, \\ E(X) &= \lambda \ \text{ and } \ Var(X) = \lambda. \end{aligned}$$

3. **The normal distribution (P.239-246,252-260)**

   If a *continuous* rv X follows a normal distribution with parameters μ and σ^2, denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$, its probability density function (pdf), mean and variance are respectively:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \text{ for all real } x,$$
$$E(X) = \mu \text{ and } Var(X) = \sigma^2.$$

1.3.3 Test statistic

1. A function $T = f(X_1, \ldots, X_n)$ of the sample observations $X_1, X_2, \cdots, X_n$ is called a *statistic*. If the observed values of $X_1, ..., X_n$ are $x_1, ..., x_n$, the observed value of T will be $t_0 = f(x_1, ..., x_n)$.

2. The sample mean and sample variance:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

   are statistics and the observed values of random variables:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right]$$

   which estimate the true mean μ and variance σ^2 respectively.

3. A statistic is a function of random variables and is also a random variable which varies from sample to sample. It's distribution is called the *sampling distribution* and it's standard deviation is called the *standard error*, e.g. $se(\bar{X})$.

4. Some results on mean and variance are:

If X_i are independent with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$,

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mu_i \text{ and } Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

If X_i are iid with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$,

$$E(\bar{X}) = \mu \text{ and } Var(\bar{X}) = \frac{\sigma^2}{n} \text{ (set } a_i = \frac{1}{n})$$

If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$.

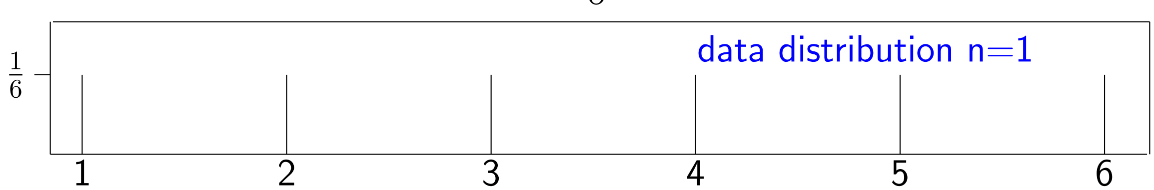
5. The Central Limit Theorem (CLT, P.246-247)

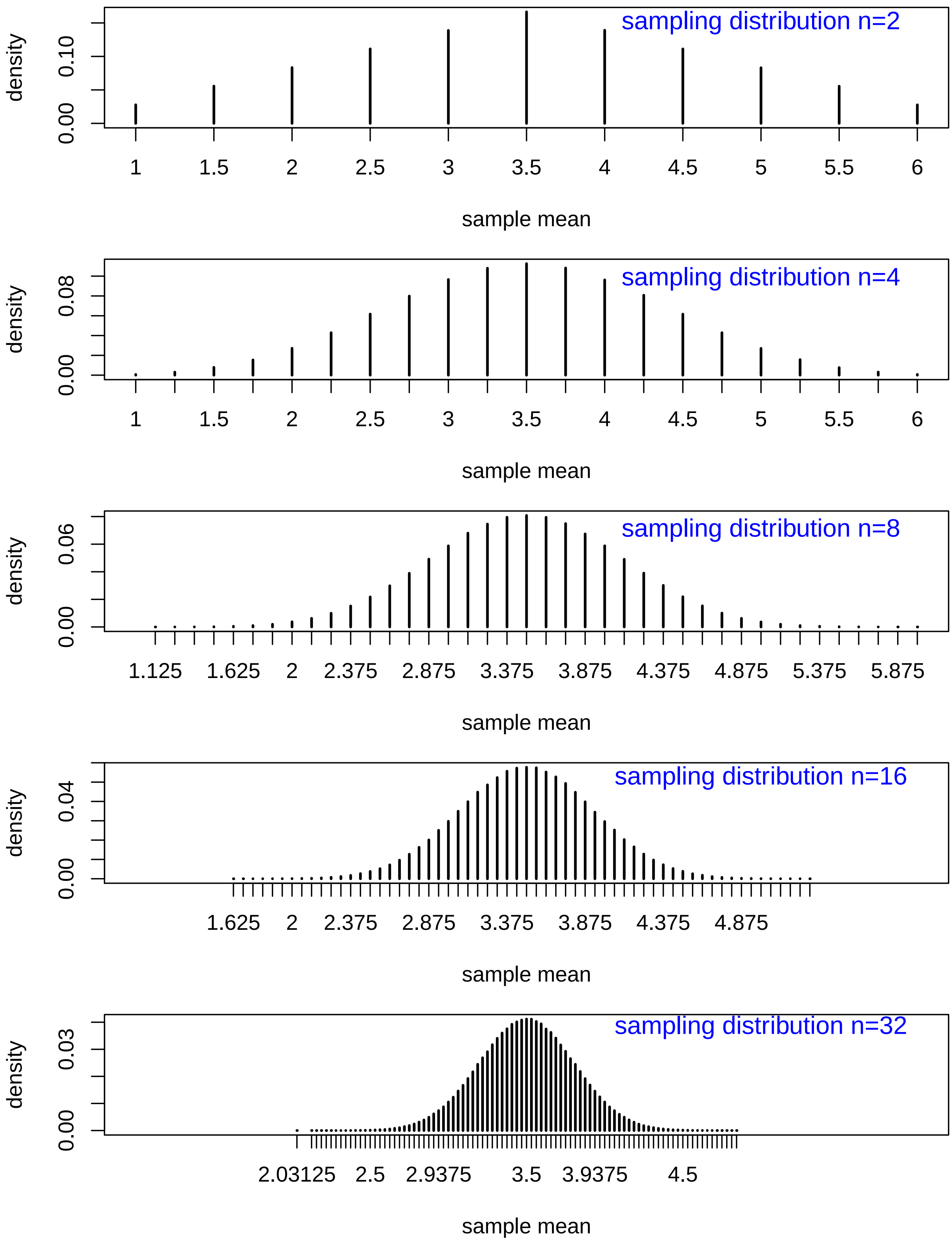
As n increases, the distribution of the sample mean $\bar{X}$ approaches normal, for any data distribution with mean μ and variance σ^2.

Hence the *standardized* variable, $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, approaches standard normal distribution $\mathcal{N}(0, 1)$ for large n.

For example, the data distribution for the outcome of a dice is *uniform* with

$$P(x = i) = \frac{1}{6}, \; i = 1, \ldots, 6.$$




sampling distribution n=2
density
0.00
0.10
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
sample mean
sampling distribution n=4
density
0.00
0.08
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
sample mean
sampling distribution n=8
density
0.00
0.06
1.125
1.625
2
2.375
2.875
3.375
3.875
4.375
4.875
5.375
5.875
sample mean
sampling distribution n=16
density
0.00
0.04
1.625
2
2.375
2.875
3.375
3.875
4.375
4.875
sample mean
sampling distribution n=32
density
0.00
0.03
2.03125
2.5
2.9375
3.5
3.9375
4.5
sample mean

Note:

(a) As the sample size n increases, the distribution of the sample mean $\bar{X}$ gradually approaches normal with the *same mean* but the *variance decreases*. These are illustrated in the diagrams.

(b) *If the data distribution for X_i is symmetric or close to normal, Z approaches $\mathcal{N}(0,1)$ at a faster rate (at lower sample size).* If X_i is not too skewed and $n = 20$ or 30, the distribution of $\bar{X}$ can be assumed normal.

This remarkable result as illustrated from the dice example is known as *The Central Limit Theorem* (CLT) and it plays an important role in statistics.

6. The statistic used in a statistical test is called the *test statistic* and it depends on

   (a) **the level of measurement**

   The test statistic for the count of successes and the measurement over a continuous range with $X_i \sim \mathcal{N}(\mu, \sigma^2)$ are respectively:

   $$\text{For a binary (class) variable,} \quad T_b = \sum_i^n X_i \sim \mathcal{B}(n, p).$$

   $$\text{For a continuous variable,} \quad T_c = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

   When the true variance σ^2 is *unknown* and is estimated by the sample variance S^2,

$$T_c = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

(b) **the number of samples**

Based on 2 normal samples $\{X_i\}$ and $\{Y_i\}$ of sizes n_x and n_y respectively,

For a matched pair sample, $$T = \frac{\overline{X} - \overline{Y}}{S_d/\sqrt{n}} \sim t_{n-1}$$

For two independent samples, $$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x+n_y-2}$$

where

S_d^2 is the *sample variance* for $n(= n_x = n_y)$ differences $D_i = Y_i - X_i$ from a *matched pair* sample $\{(X_i, Y_i)\}$ and

S_p^2 is the *pooled variance* estimate of S_x^2 and S_y^2 for the *two independent* samples $\{X_i\}$ amd $\{Y_i\}$.

(c) **the variable of interest**

Based on a normal sample $\{X_i\}$,

For testing the mean, $$T = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

For testing the variance, $$T = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}.$$

(d) **the distribution assumption**

For parametric test, we assume that $X_1, X_2, ..., X_n$ follow certain distribution with a distribution function $F_\theta(x) = \Pr(X < x)$ depending on certain unknown parameter θ, e.g. μ, σ^2 or p, which is to be tested.

For non-parametric test, we only assume that $X_1, X_2, ..., X_n$ are iid and not too skewed. We do not specify a particular type of distribution for X_i. If $X_1, X_2, ..., X_n$ are very skewed, we transform the data to make it more symmetric. *Skewness* can be checked by examining the *box-plot* or *stem-and-leaf* plot.

2 One sample t-test

2.1 Steps of a general test (P.350-356)

1. **Hypotheses**

   The statement against which you *search for evidence* is called the *null hypothesis*, and is denoted by H_0. It is generally a "no difference" statement.

   The *statement you claim* is called the *alternative hypothesis*, and is denoted by H_1. We generally choose

   $$H_0: \ \theta = \theta_0; \ \text{against} \ H_1: \begin{array}{l} \theta > \theta_0 \text{ (upper-side alternative)}, \\ \theta < \theta_0 \text{ (lower-side alternative)}, \\ \theta \neq \theta_0 \text{ (two-sided alternative)} \end{array}$$

   In the flu vaccine example, we assume that the number X of people who receive the new vaccine and do not contact the virus in the second year follows $\mathcal{B}(n, p)$, then the hypotheses are:

   $$H_0: \ p = 0.20, \quad H_1: \ p > 0.20.$$

   In the beer content example, we assume that the beer content $X \sim \mathcal{N}(\mu, \sigma^2)$, then the hypotheses are:

   $$H_0: \ \mu = 375, \quad H_1: \ \mu < 375.$$

   In the height comparison example, we assume that the height of girls $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, and the height of boys $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then the hypotheses are:

   $$H_0: \ \mu_1 - \mu_2 = 0, \quad H_1: \ \mu_1 - \mu_2 > 0.$$

   In the body temperature example, the hypotheses are:

   H_0: the measurements follow a normal distribution.

   H_1: the measurements do not follow a normal distribution.

2. **Assumptions**

   Each data $X_1, X_2, ..., X_n$ is *chosen at random* from a population and hence is an *iid* rv from the *same* population distribution.

3. **Test statistic**

   Since observations X_i vary from sample to sample we can never be sure whether H_0 is true or not. We use a *test statistic* $T = f(X_1, ..., X_n)$ to test if the data are consistent with H_0 such that

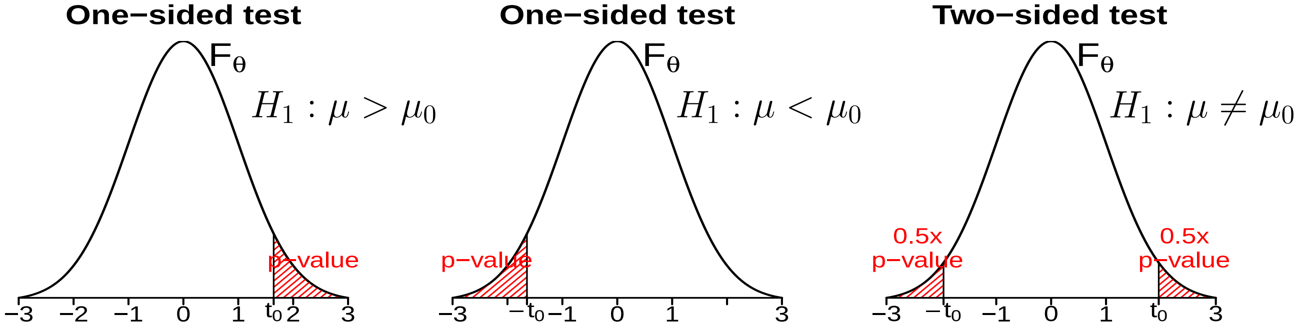
   1. the distribution of T is known assuming H_0 is true;
   2. the *large* (positive or negative depending on H_1) observed value of T is taken as evidence of poor agreement with H_0.

4. **P-value**

   The p-value (*observed significance level*) is defined as

$$
\begin{aligned}
p\text{-value} &= 1 - F_\theta(t_0) = \Pr(T \geq t_0 \,|H_0) & H_1: \ \theta > \theta_0 \text{ or}\\
p\text{-value} &= F_\theta(-t_0) = \Pr(T \leq -t_0 \,|H_0) & H_1: \ \theta < \theta_0 \text{ or}\\
p\text{-value} &= 2F_\theta(-|t_0|) = 2\Pr(T \geq |t_0| \,|H_0) & H_1: \ \theta \neq \theta_0
\end{aligned}
$$

   which is the probability of getting the observed test statistic t_0 and *more extreme* values *assuming that H_0 is true*.



The p-value is represented by the area under the density function from the distribution function F_θ.

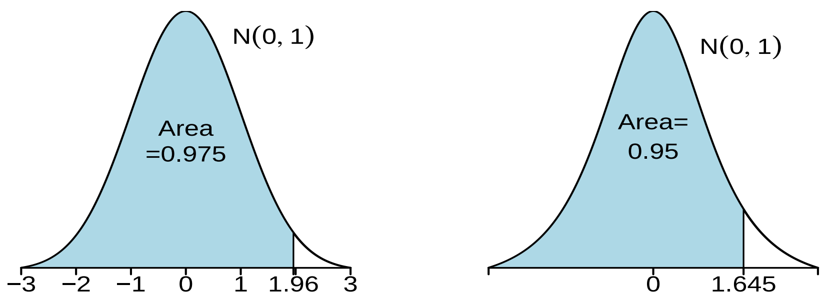
In R,

the codes for calculating various probabilities and quantiles are:

For normal distribution:

`pnorm(1.96)` gives $\Phi(1.96) = \Pr(z < 1.96) = 0.9750$.
`qnorm(0.95)` gives $c = \Phi^{-1}(0.95) = 1.6449$ if $\Pr(z < c) = 0.95$.

Note: $\Phi(x)$ denotes the distribution function (df) F for normal distribution and it gives the lower area or probability up to x.

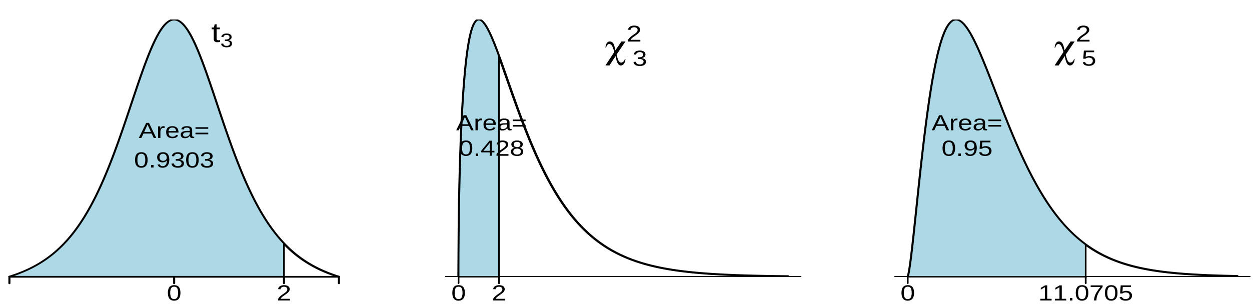


Similarly for t and χ^2 distributions,

`pt(2,3)` gives $\Pr(t_3 \le 2) = 0.930337$,

`pchisq(2,3)` gives $\Pr(\chi^2_3 \le 2) = 0.4275933$ and

`qchisq(0.95,5)` gives 11.0705, i.e., $\Pr(\chi^2_5 \le 11.0705) = 0.95$.



5. **Decision on the null hypothesis**

   Observed *large* positive or negative value of t_0 and hence small value of p-value is taken as evidence of poor agreement with H_0.

   (a) If the p-value is small, then *either* H_0 is true and the poor agreement is due to an unlikely event, *or* H_0 is false. Therefore,

   The smaller the p-value, the stronger the evidence against H_0 in favour of H_1.

   (b) Large p-value does not mean that there is evidence that H_0 is true, but only that the test *detects no inconsistency* between the claim on H_0 and the results of the experiment.

The level of *significance*, α indicates the amount of the evidence provided by data against H_0 and is interpreted as (if $\alpha = 0.05$)

p-value > 0.10	The data are consistent with H_0 (accept H_0).
p-value $\in (0.05, 0.10)$	Borderline evidence against H_0 (accept H_0).
p-value $\in (0.025, 0.05)$	Reasonably strong evidence against H_0 (reject H_0).
p-value $\in (0.01, 0.025)$	Strong evidence against H_0 (reject H_0).
p-value < 0.01	Very strong evidence against H_0 (reject H_0).

2.2 One sample t-test for mean in normal population (P.385-388,394-405)

Example: (Beer contents) Beer contents in a pack of six bottles (in millilitres) are:

374.8, 375.0, 375.3, 374.8, 374.4, 374.9.

Is the mean beer content is less than 375 mL as claimed on the label?

How to test for the mean from one sample of observations measured over a continuous range?

Suppose we have a sample $X_1, X_2, ..., X_n$ of the size n drawn from a normal population with an unknown variance σ^2. Let $x_1, x_2, ..., x_n$ be the observed values. We want to test the population mean μ.

1. **Hypothesis:** $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$, $\mu < \mu_0$, $\mu \neq \mu_0$.
2. **Test statistic:** $t_0 = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$, under H_0
3. **Assumptions:** X_i are iid rv and follow $\mathcal{N}(\mu, \sigma^2)$ with σ^2 unknown if n is small. No assumption for X_i if n is large.
4. P**-value:**

$$\begin{array}{rl} \Pr(t_{n-1} \geq t_0) & \text{for } H_1: \ \mu > \mu_0; \\ \Pr(t_{n-1} \leq -t_0) & \text{for } H_1: \ \mu < \mu_0; \\ 2\Pr\left(t_{n-1} \geq |t_0|\right) & \text{for } H_1: \ \mu \neq \mu_0. \end{array}$$

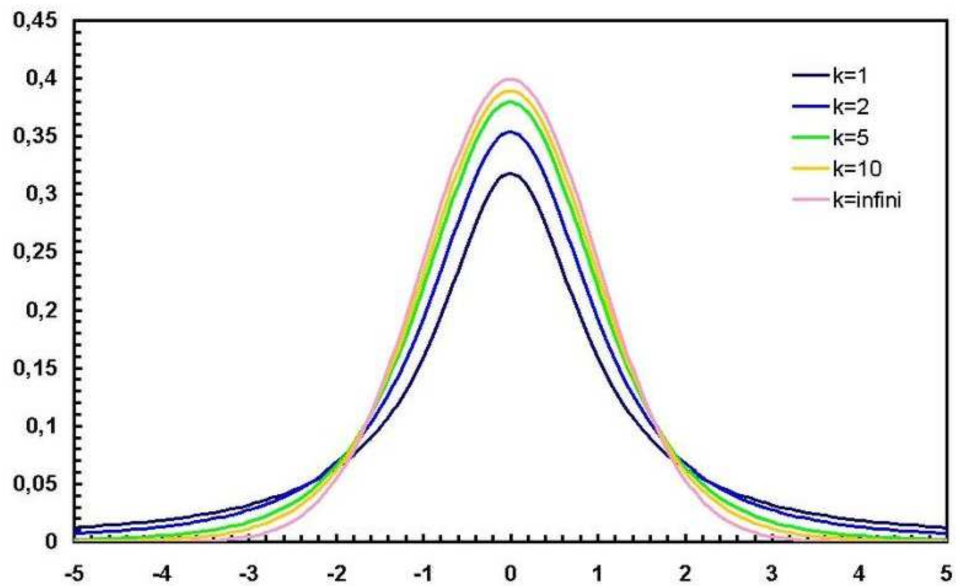
   In R, `pt(t,n-1)` gives $\Pr(t_{n-1} \leq t)$.

5. **Decision:** Reject H_0 in favour of H_1 if the p-value is small.

Remarks:

1. For small sample ($n < 20$), the t-test is sensitive to the normality assumption. To check this assumption, the points in the qq-plot should lie closely to a *straight line*. The R commands are `qqnorm` and `qqline`. However the plots may be misguided if $n \leq 10$.

2. For large sample ($n \geq 20$), CLT assures that t-test can be used if the data come from a *random* design. The t distribution approaches standard normal distribution as df $\to \infty$, i.e. $\Pr(t_{n-1} \leq t_0) \approx \Pr(Z < t_0) = \Phi(t_0)$. If the data are too skew as shown in the qqplot with large *outliers* (points lie far away from the line) on one side, transformation is needed.

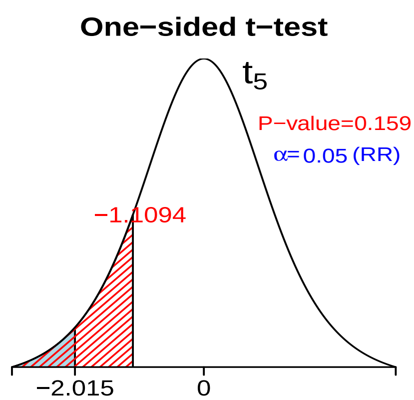


Density functions of Student's t distribution with k degrees of freedom.

Example: (Beer contents)

Solution: We have $n = 6$, $\bar{x} = 374.87$ and $s^2 = 0.087$. The one sample t-test of the mean content is

1. **Hypothesis:** $H_0 : \mu = 375 \quad \text{vs} \quad H_1 : \mu < 375.$
2. **Test statistic:**
$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{374.87 - 375}{\sqrt{0.087}/\sqrt{6}} = -1.1094.$$
Large negative value of t_0 will argue against H_0 in favour of H_1.
3. **Assumption:** As the sample size $n = 6$ is very small, we assume that the beer contents $X_i \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is unknown. Then $t_0 \sim t_{n-1}$.
4. P**-value**:
$\Pr(t_5 \leq -1.1094) = 0.1589$ (from t-table or `pt(-1.1094,5)` in R)
5. **Decision**: Since the p-value > 0.05, the data is consistent with the claim on H_0 that the mean content is 375 mL.



In R

```
> x=c(374.8, 375.0, 375.3, 374.8, 374.4, 374.9)
> t.test(x, alternative="less", mu=375)
```

```
        One Sample t-test

data:  x
t = -1.1094, df = 5, p-value = 0.1589
alternative hypothesis: true mean is less than 375
95 percent confidence interval:
     -Inf 375.1088
sample estimates:
mean of x
 374.8667

> n=length(x)  # for checking
> mu0=375
> xbar=mean(x)
> s=sd(x)
> t0=(xbar-mu0)/(s/sqrt(n))
> p.value=pt(t0,n-1)
> c(n,xbar,s,t0,p.value)
[1]   6.0000000 374.8666667   0.2943920  -1.1094004
      0.1588721
```

As the hypothesized mean $\mu_0 = 375$ lies *inside*

$$\text{the 95\% CI for } \mu = (-\infty, 375.1088)$$

which indicates *no significant difference* between μ_0 and μ, H_0 is accepted.

Note that qq-plot is not used to check for normality because the sample size of $n = 6$ is too small to give a reliable result.

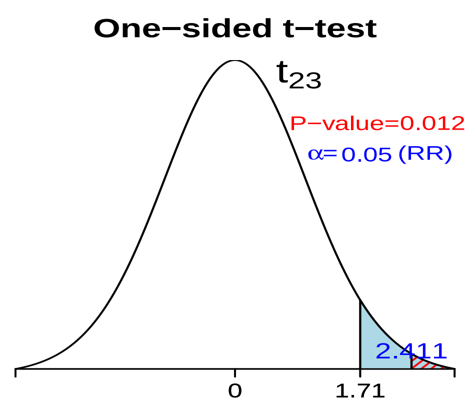
Example: (Sales contacts) A vice president in charge of sales for a large corporation claims that salespeople are averaging more than 15 sales contacts each week. As a check on his claim, 24 salespeople are selected at random, and the number of the contacts made by each is recorded for a single randomly selected week as follows:

```
18, 17, 18, 13, 15, 16, 21, 14, 24, 12, 19, 18,
17, 16, 15, 14, 17, 18, 19, 20, 13, 14, 12, 15
```

Does the evidence contradict the vice president's claim?

Solution: The one sample t-test for the mean of sales contacts is

1. **Hypotheses:** $H_0 : \mu = 15$ vs $H_1 : \mu > 15$
2. **Test statistic:** We have $n = 24$, $\bar{x} = 16.4583$ and $s = 2.9632$.
$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{16.4583 - 15}{2.9632/\sqrt{24}} = 2.4110.$$
Large value of t_0 will argue against H_0 in favour of H_1.
3. **Assumption:** No particular assumption on X_i. As $n = 24 > 20$, $t_0 \sim t_{23}$ by CLT.
4. **P-value:** p-value $= \Pr(t_{23} \geq 2.4110) = 0.0121$
5. **Decision:** Since the p-value < 0.05, there is strong evidence against H_0. The average number of sales contacts per week exceeds 15.



In R,

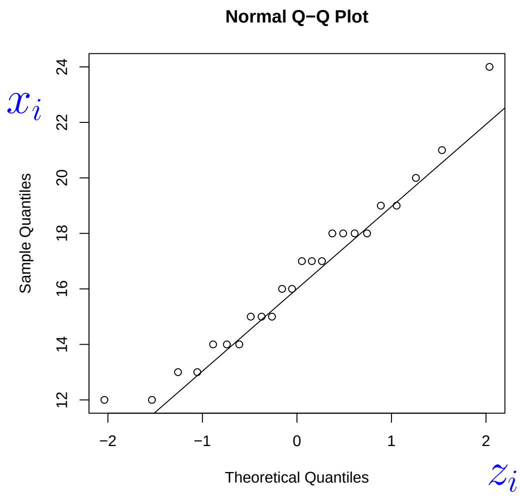
```
> x<-c(18,17,18,13,15,16,21,14,24,12,19,18,17,16,15,14,17,
       18,19,20,13,14,12,15)
> n=length(x)
> mu0=15
> qqnorm(x)
> qqline(x)
> t.test(x, alternative="greater",mu=15)

        One Sample t-test

data:  x
t = 2.411, df = 23, p-value = 0.01215
alternative hypothesis: true mean is greater than 15
95 percent confidence interval:
 15.42167      Inf
sample estimates:
mean of x
 16.45833

> barx=mean(x)  #for checking
> s=sqrt(var(x))
> t0=(barx - mu0)/(s/sqrt(n))
> p=pt(t0,n-1,lower.tail=F)
> c(barx,s,t0,p)
[1] 16.45833333  2.96324098  2.41099024  0.01214894
```

As the hypothesized mean $\mu_0 = 15$ lies *outside*

$$\text{the 95\% CI for } \mu = (15.42167, \infty)$$

which indicates a *significant difference* between μ_0 and μ, H_0 is rejected.

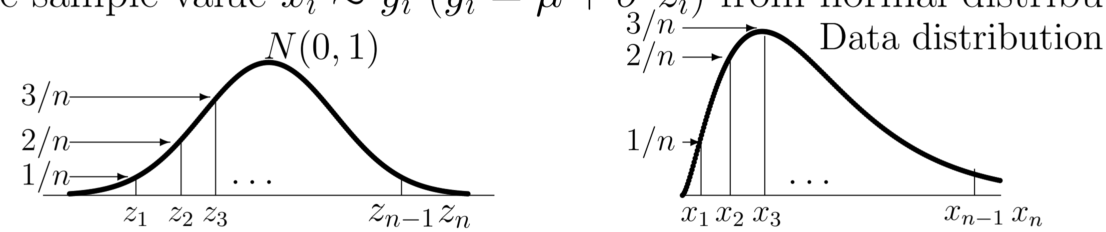
Note: the normality assumption for the data cannot be closely satisfied since there are outliers in the qq-plot but CLT assures normality for the sample mean as the sample size $n = 24$ is relatively large.



Remarks:

1. The normal qq-plot using R commands `qqnorm(d)` test for *normality* assumption.

   The qq-plot plots sample quantiles x_i against standard normal quantiles z_i, i.e. $(\frac{y_i-\hat{\mu}}{\hat{\sigma}}, x_i)$ when lower area is $\frac{i}{n}$. If a point lies on the line, the sample value $x_i \approx y_i$ $(y_i = \hat{\mu} + \hat{\sigma}\ z_i)$ from normal distribution.



2. For small sample size $n < 20$, the points in qq-plot should be nearly along a straight line.

3. For large sample size $n \geq 20$, the t-test can be used if the data X_i are not too skew. It is not necessary to assume that the sample is drawn from a normal population because CLT assures normality of sample mean $\bar{x}$ if n is large.

3 Paired sample t-test and Z-test

What if the sample is a paired sample?

3.1 Paired sample t-test for mean

Example: (Smoking) Blood samples from 11 individuals before and after they smoked a cigarette are used to measure aggregation of blood platelets.

Before:	25	25	27	44	30	67	53	53	52	60	28;
After:	27	29	37	36	46	82	57	80	61	59	43.

Is the aggregation affected by smoking?

Solution: Let X_i and Y_i be the aggregation of blood platelets before and after smoking by the same individual.

```
-2, -4, -10, 8, -16, -15, -4, -27, -9, 1, -15
```

The paired sample t-test for the difference $D_i = X_i - Y_i$ is:

1. **Hypotheses:** $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x > \mu_y,\ \mu_x < \mu_y,\ \mu_x \neq \mu_y$

   Or $H_0 : \mu_d = 0$ vs $H_1 : \mu_d > 0,\ \mu_d < 0,\ \mu_d \neq 0$

   where $\mu_d = \mu_x - \mu_y$.

2. **Test statistic:** $$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

   where $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ is the sample mean and

   $s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$ is the sample variance.

3. **Assumption:** The differences $d_j \sim N(\mu, \sigma^2)$, where σ^2 is unknown if n is small. No distribution assumption for d_i if n is large. Then $t_0 \sim t_{n-1}$ under H_0.

4. P**-value:**
$$\begin{aligned} &\Pr\left(t_{n-1} \geq t_0\right) && \text{for } H_1 : \mu_d > 0, \\ &\Pr\left(t_{n-1} \leq -t_0\right) && \text{for } H_1 : \mu_d < 0, \\ &2\Pr\left(t_{n-1} \geq |t_0|\right) && \text{for } H_1 : \mu_d \neq 0, \end{aligned}$$

5. **Decision:** Reject H_0 in favor of H_1 if the p-value is small.

Note that large positive or negative value of t_0 will argue against H_0 in favor of H_1.

Hence the paired sample t-test on whether the aggregation is affected by smoking is

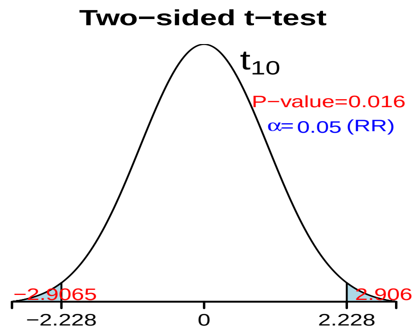
1. **Hypotheses:** $H_0 : \mu_d = 0 \quad \text{vs} \quad H_1 : \mu_d \neq 0.$

2. **Test statistic:** We have $n = 11$, $\bar{d} = -8.4545$ and $s_d = 9.6474$.
$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-8.4545}{9.6474/\sqrt{11}} = -2.9065.$$
Large value of t_0 will argue against H_0 in favor of H_1.

3. **Assumption:** Since $n = 11$ is small, we assume that the differences $d_i \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown. Hence $t_0 \sim t_{10}$. The normal QQ plot shows that the normality assumption is satisfied.

4. P**-value:** $p\text{-value} = 2\Pr(t_{10} < -2.9065) = 0.0157$

5. **Decision:** Since the p-value < 0.05, there is strong evidence against H_0. The aggregation is affected by smoking.



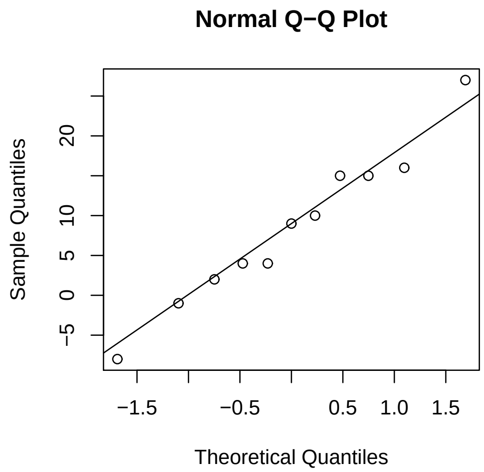
In R,

```
> x<- c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)
> y<- c(27, 29, 37, 36, 46, 82, 57, 80, 61, 59, 43)
> d=y-x
> d
 [1]  2  4 10 -8 16 15  4 27  9 -1 15
> boxplot(d)
> title("boxplot of diff")
> qqnorm(d)
> qqline(d)
> t.test(x,y, alternative="two.sided", mu=0, paired=T)

        Paired t-test

data:  x and y
t = -2.9065, df = 10, p-value = 0.01566
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.93577  -1.97332
sample estimates:
mean of the differences
              -8.454545
```

Note:

1. $d_0 = 0$ does not lie in $(-14.93577, -1.97332)$, the 95% CI for D. Hence d_0 is significantly difference from $\bar{D}$ and we reject H_0.

2. The assumption of normal data distribution is satisfied as the points in the qq plot are closed to the qq line.



3.2 One sample Z-test (P.357-360)

Example: (Birth weights) The birth weights of a random sample of 14 boys born to mothers who smoked heavily during pregnancy were recorded (in ounces). The data are:

$$79,\ 92,\ 88,\ 98,\ 109,\ 109,\ 112,$$
$$88,\ 105,\ 89,\ 121,\ 71,\ 110,\ 96.$$

From record, the population standard deviation is known to be $\sigma = 15$ ounces. Is it reasonable to assume that on average, boys born to mothers who smoke have lower birthweights than the national average of 109 ounces (3.09kg)?

What if the population sd is know? Should we use it in the calculation?

We want to test a population mean μ based on a random sample $X_1, X_2, ..., X_n$ drawn from $\mathcal{N}(\mu, \sigma^2)$ where the population variance σ^2 is *known*.

We use the z-test to test hypothesis about the population mean μ.

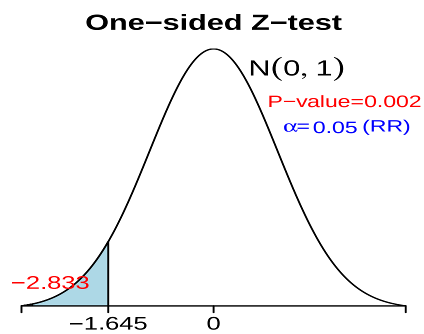
1. **Hypothesis:** $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$, $\mu < \mu_0$, $\mu \neq \mu_0$.
2. **Test statistic:** $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
3. **Assumption:** If n is small, $X_i \sim \mathcal{N}(\mu, \sigma^2)$. If n is large, no distribution assumption is needed for X_i. In both cases, σ^2 is known. Then $z_0 \sim \mathcal{N}(0, 1)$.
4. ***p*-value:** $\Pr(Z > z_0) = 1 - \Phi(z_0)$ for $H_1 : \mu > \mu_0$,
   $\Pr(Z < -z_0) = \Phi(-z_0)$ for $H_1 : \mu < \mu_0$;
   $2\Pr(Z > |z_0|) = 2\big(1 - \Phi(|z_0|)\big)$ for $H_1 : \mu \neq \mu_0$.
5. **Decision:** Reject H_0 in favor of H_1 if the p-value is small.

Note that the paired sample z-test may be performed similarly.

Example: (Birth weights)

Solution: Let μ denote the mean birth weight for boys born to mothers who smoke. The qq-plot below shows that the normality assumption holds. The *z-test* for the *mean birth weight* is

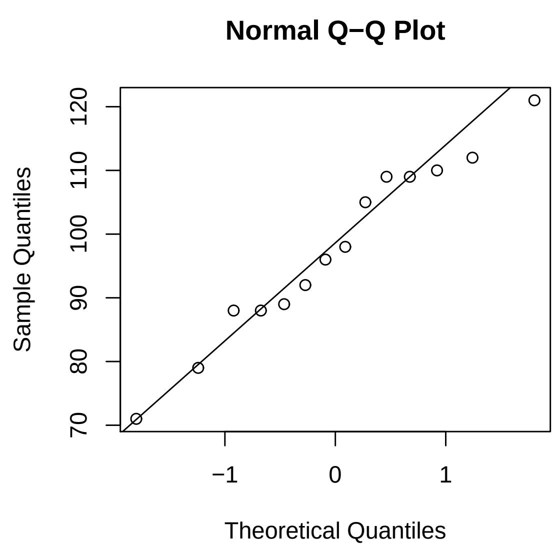
1. **Hypotheses:** $H_0 : \mu = 109$ against $H_1 : \mu < 109$.
2. **Test statistic:** $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{97.6429 - 109}{15/\sqrt{14}} = -2.8330.$
3. **Assumption:** Since $n = 14$ is small, we assume that the birth-weight $X_i \sim \mathcal{N}(109, 15^2)$. Then $z_0 \sim \mathcal{N}(0, 1)$ under H_0. The normal qq-plot shows that the normality assumption is satisfied.
4. **P-value:** p-value $= P(Z < -2.8330) = 1 - 0.9977 = 0.0023$.
5. **Decision:** Since P-value is < 0.05, we reject H_0 and conclude that there is strong evidence against H_0. Boys born to mothers who smoke have lower birthweights.



In R,

```
> x=c(79,92,88,98,109,109,112,88,105,89,121,71,110,96)
> n=length(x)
> mu0=109
```

```
> xbar=mean(x)
> sd0=15
> z0=(xbar-mu0)/(sd0/sqrt(n))
> p.value=pnorm(z0)
> c(n,xbar,sd0,z0,p.value)
[1] 14.00000 97.642857143 15.0000 -2.832969164  0.002305892
> boxplot(x)
> title("Birthweights")
> qqplot(x)
> qqline(x)
```



Since the points in the qq-plot lie reasonably close to the qq line, the assumption of normal data distribution is approximately satisfied.

4 Critical value and rejection region

Apart from using p-value, is there other ways to make decision?

4.1 Critical value for decision rule

To test the hypotheses, our decision rule is to reject H_0

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu > \mu_0, \text{ or...}$$

when the p-value is less than certain fixed preassigned levels, say $\alpha = 0.05, 0.10$, etc. In other words, we accept or reject H_0 according to $p > \alpha$ or $p \leq \alpha$. The α is called the *significance level* of the test, which is the boundary between accept and reject H_0.

We may also find a *critical value* $t_{n-1}(\alpha)$ at the significance level α for the test statistic $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ such that

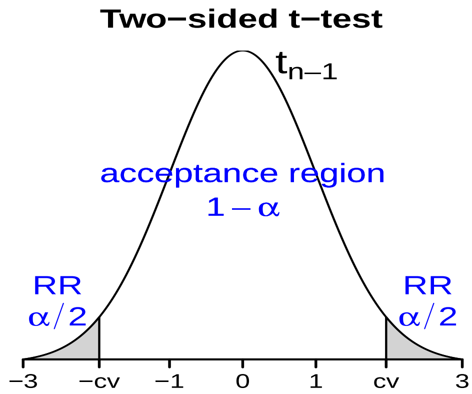
$$\Pr(t_{n-1} \geq t_{n-1}(\alpha) | H_0) = \alpha, \quad \text{or ...}$$

The critical value depends on both t_{n-1}, the distribution of T under H_0, and α. *Decision rule* at level α can be defined as

reject H_0 if $t_0 \geq t_{n-1}(\alpha)$,
accept H_0 if $t_0 < t_{n-1}(\alpha)$.

Alternatively, for the test statistic $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, the decision rule is

reject H_0 if $z_0 \geq z(\alpha)$,
accept H_0 if $z_0 < z(\alpha)$.



Rejection region for two sided test
For one-sided test, $\alpha/2$ from either side should be changed to α.

From the t table,
With $\alpha = 0.1$, $t_3(\alpha) = 1.638$, $t_5(\alpha) = 1.476$, ...
With $\alpha = 0.05$, $t_3(\alpha) = 2.353$, $t_5(\alpha) = 2.015$, ...

From the standard normal table,
$z(0.1) = 1.28$, $z(0.05) = 1.645$, $z(0.025) = 1.96$, $z(0.02) = 2.054$,
$z(0.01) = 2.326$, $z(0.005) = 2.576$, ...

In R, the critical values are `qt(p,n-1)`, `qnorm(p)`, `qchisq(p,n-1)`, `qf(p,m-1,n-1)` for t, normal, χ^2 and F distributions respectively.

4.2 Rejection region for test statistics

Suppose $X_i, \ldots, X_n$ are drawn from some population. Given a significance level α, we want to test on the population mean.

1. **Hypothesis:** $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0,\ \mu < \mu_0,\ \mu \neq \mu_0$
2. **Test statistic:** $t_0 = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$ or $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
3. **Assumptions:** X_i are iid rv with $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown and known respectively.
4. **Rejection region:**

$$\begin{array}{lll} t_0 \ \geq \ t_{n-1}(\alpha) & \text{or } \ z_0 \geq z(\alpha) & \text{for } \ H_1 : \ \mu > \mu_0; \\ t_0 \ \leq -t_{n-1}(\alpha) & \text{or } \ z_0 \leq \ -z(\alpha) & \text{for } \ H_1 : \ \mu < \mu_0; \\ |t_0| \geq t_{n-1}(\alpha/2) & \text{or } \ |z_0| \ \geq z(\alpha/2) & \text{for } \ H_1 : \ \mu \neq \mu_0, \\ \text{i.e. } t_0 \leq -t_{n-1}(\alpha/2) & \text{or } \quad z_0 \leq -z(\alpha/2) & \\ \ \text{or } t_0 \geq t_{n-1}(\alpha/2)\} & \quad \text{or } z_0 \geq z(\alpha/2)\} & \end{array}$$

   where $t_{n-1}(\beta)$ or $z(\beta)$, $\beta = \alpha$ or $\alpha/2$ is the critical value given by

$$\Pr(t_{n-1} \geq t_{n-1}(\beta)) = \beta \text{ or } \Pr(Z \geq z(\beta)) = \beta.$$

5. **Decision:** We reject H_0 if t_0 or $z_0 \in$ RR.

The complement of the rejection region is called *acceptance region*.

If the sample size is large enough ($n \geq 20$), the rejection regions of the large sample t-test and z-test are nearly same.

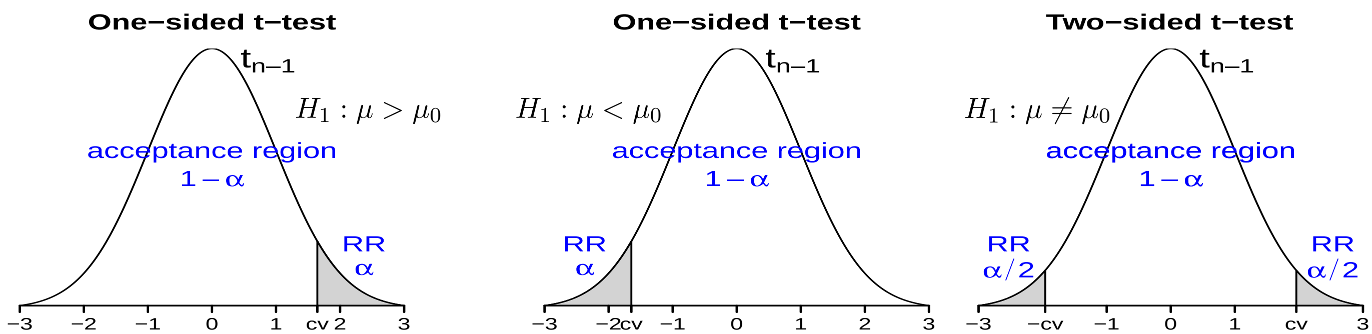
4.3 Rejection region for sample mean

The rejection regions for the test using test statistic $t_0 = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} \geq t_{n-1}(\alpha)$ or $z_0 = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq z(\alpha)$ on the standardized scale can be transformed to measurement scale :

$\{\bar{x} : \bar{x} \geq k_0 = \mu_0 + t_{n-1}(\alpha)s/\sqrt{n}\}$, or $\{\bar{x} : \bar{x} \geq k_0 = \mu_0 + z(\alpha)\sigma/\sqrt{n}\}$ for $H_1 : \ \mu > \mu_0$;

$\{\bar{x} : \bar{x} \leq k_0 = \mu_0 - t_{n-1}(\alpha)s/\sqrt{n}\}$, or $\{\bar{x} : \bar{x} \leq k_0 = \mu_0 - z(\alpha)\sigma/\sqrt{n}\}$ for $H_1 : \ \mu < \mu_0$;

$\{\bar{x} : \bar{x} \leq k_0 = \mu_0 - t_{n-1}(\frac{\alpha}{2})s/\sqrt{n}$ or $\bar{x} \geq k_0 = \mu_0 + t_{n-1}(\frac{\alpha}{2})s/\sqrt{n}\}$, or $\{\bar{x} : \bar{x} \leq k_0 = \mu_0 - z(\frac{\alpha}{2})\sigma/\sqrt{n}$ or $\bar{x} \geq k_0 = \mu_0 + z(\frac{\alpha}{2})\sigma/\sqrt{n}\}$ for $H_1 : \ \mu \neq \mu_0$;



standardized scale: 0 $\quad t_{n-1}(\alpha)$ $\quad -t_{n-1}(\alpha)$ $\quad 0$ $\quad -t_{n-1}(\frac{\alpha}{2})$ $\quad 0$ $\quad t_{n-1}(\frac{\alpha}{2})$

measurement scale: μ_0 $\quad \mu_0{+}t_{n-1}(\alpha)\frac{s}{\sqrt{n}}$ $\quad \mu_0{-}t_{n-1}(\alpha)\frac{s}{\sqrt{n}}$ $\quad \mu_0$ $\quad (\mu_0{-}t_{n-1}(\frac{\alpha}{2})\frac{s}{\sqrt{n}}$ $\quad \mu_0$ $\quad \mu_0{+}t_{n-1}(\frac{\alpha}{2})\frac{s}{\sqrt{n}})$

Acceptance and rejection regions for t-test under the three types of H_1.

Similarly when the distribution is $\mathcal{N}(0,1)$, replace $t_{n-1}(\cdot)$ by $z(\cdot)$ and s by σ.

Example: For the above four examples, find the rejection regions on the standardized and measurement scales.

Solution:

1. (Beer contents) with $n = 6$, $\bar{x} = 374.87$, $s^2 = 0.087$, $t_0 = -1.1094$, $H_1: \ \mu < 375$ and not rej. H_0.

3". **Test statistic:** $\bar{x} = 374.87$

4'. **Rejection region:** $t_0 < -t_5(0.05) = -2.015$

or 4". **Rejection region:**
$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -t_5(0.05)$$
$$\bar{x} < \mu_0 - t_{n-1}(0.05)\, s/\sqrt{n}$$
$$\text{i.e. } \bar{x} < 375 - 2.015\sqrt{0.087}/\sqrt{6}$$
$$\text{i.e. } \bar{x} < 374.7574$$

5'. **Decision:** Since $t_0 = -1.1094 > -2.015$, accept H_0.

or 5". **Decision:** $\bar{x} = 374.87 > 374.76$, accept H_0.

2. (Sales contacts) with $n = 24$, $\bar{x} = 16.4583$, $s = 2.9632$, $t_0 = 2.4110$, $H_1: \ \mu > 15$ and reject H_0.

3". **Test statistic:** $\bar{x} = 16.4583$

4'. **Rejection region:** $t_0 > t_{23}(0.05) = 1.714$

or 4". **Rejection region:**
$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{23}(0.05)$$
$$\bar{x} > \mu_0 + t_{n-1}(0.05)\, s/\sqrt{n}$$
$$\text{i.e. } \bar{x} > 15 + 1.714 \cdot 2.9632/\sqrt{24}$$
$$\text{i.e. } \bar{x} > 16.0367$$

5'. **Decision:** Since $t_0 = 2.4110 > 1.714$, reject H_0.

or 5". **Decision:** Since $\bar{x} = 16.46 > 16.04$, reject H_0.

3. (Smoking) with $n = 11$, $\bar{d} = -8.4545$, $s_d = 9.6474$, $t_0 = -2.9065$, $H_1 : \mu_d \neq 0$ and reject H_0.

3". **Test statistic:** $\bar{d} = -8.4545$

4'. **Rejection region:** $|t_0| > t_{10}(0.025) = 2.228$

(for test statistics) i.e. $t_0 > 2.228$ or $t_0 < 2.228$

or 4". **Rejection region:** $\dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}} < -t_{10}(0.025) = -2.228$

(for sample mean)

$$\bar{d} < \mu_d - t_{n-1}(0.025)\, s_d/\sqrt{n}$$

$$\text{i.e. } \bar{d} < 0 - 2.228 \cdot 9.6474/\sqrt{11}$$

$$\text{i.e. } \bar{d} < -6.4808$$

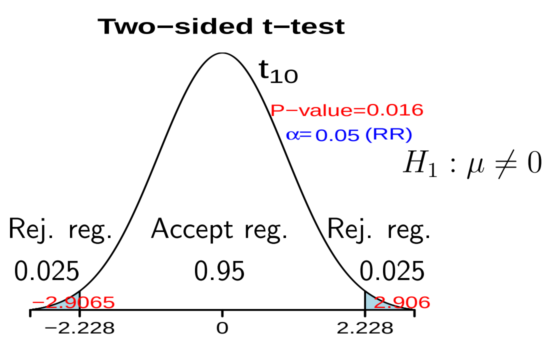
$$\text{or } \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} > t_{10}(0.025) = 2.228$$

$$\bar{d} > \mu_d + t_{n-1}(0.025)\, s_d/\sqrt{n}$$

$$\text{i.e. } \bar{d} > 0 + 2.228 \cdot 9.6474/\sqrt{11}$$

$$\text{i.e. } \bar{d} > 6.4808$$

5". **Decision:** Since $t_0 = -2.91 < -2.228$ or $\bar{x} = -8.45 < -6.48.76$, reject H_0



	RR		AR	RR	
Standardized scale.	-2.91 t_0	-2.23 $-t_{10,\frac{\alpha}{2}}$	0	2.23 $t_{10,\frac{\alpha}{2}}$	$t_0 = -2.91 < -2.23$. In RR. Rej. H_0.
Measurement scale.	-8.45 $\bar{x}$	-6.48 $\mu_0 - t\frac{s}{\sqrt{n}}$	0 μ_0	6.48 $\mu_0 + t\frac{s}{\sqrt{n}}$	$\bar{d} = -8.45 < -6.48$. In RR. Rej. H_0.

4. (Birth weights) with $n = 14$, $\bar{x} = 97.6429$, $\sigma = 15$, $z_0 = -2.8330$, $H_1: \ \mu < 109$ and reject H_0.

3". **Test statistic:** $\bar{x} = 97.6429$

4'. **Rejection region:** (for test statistics) $z_0 < -z(0.05) = -1.645$

or 4". **Rejection region:** (for sample mean)

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z(0.05)$$

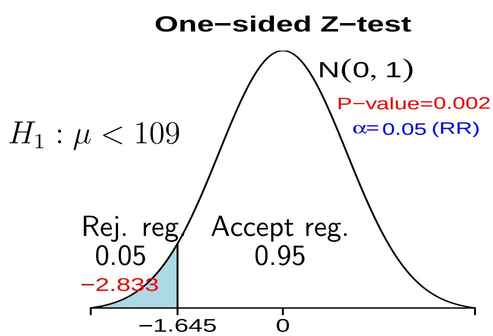
$$\bar{x} < \mu_0 - z(\alpha)\sigma/\sqrt{n}$$

i.e. $\bar{x} < 109 - 1.645 \cdot 15\sqrt{14}$

i.e. $\bar{x} < 102.4053$

5'. **Decision:** Since $z_0 = -2.8330 < -1.645$, reject H_0.

or 5". **Decision:** Since $\bar{x} = 97.64 < 102.41$, reject H_0



	RR	AR	
Standardized scale.	-2.83 (z_0)	-1.65 ($-z_\alpha$)	0	$t_0 = -2.83 < -1.65$. In RR. Rej. H_0.
Measurement scale.	97.6 ($\bar{x}$)	102.4 ($\mu_0 - z\frac{\sigma}{\sqrt{n}}$)	109 (μ_0)	$\bar{x} = 97.6 < 102.4$. In RR. Rej. H_0.

R for the smoking example,

```
> x<- c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)
> y<- c(27, 29, 37, 36, 46, 82, 57, 80, 61, 59, 43)
> d=y-x
> d
 [1]  2  4 10 -8 16 15  4 27  9 -1 15
```

```
> sdd = sd(d)
> sdd
[1] 9.64742
> n=length(d)
> mu0=0
> cv=qt(0.975,n-1)
> cv
[1] 2.228139
> rrlower=mu0-qt(0.975,n-1)*sdd/sqrt(n)
> rrupper=mu0+qt(0.975,n-1)*sdd/sqrt(n)
> c(rrlower,rrupper)
[1] -6.481225  6.481225
```

5 Power and sample size

5.1 Type of error in drawing decision (P.366-369)

If we reject H_0, we have at most α chance of wrongly reject H_0. What if we accept H_0?

Decision rule for statistical tests with fixed level α is:

> if the test statistic falls in the rejection region, we reject H_0, and accept H_1.

Then two types of errors can be made in reaching a decision for the hypotheses:

$$H_0: \ \mu = \mu_0 \text{ vs } H_1: \ \mu > \mu_0$$

1. *Type I* error: reject H_0, but in fact H_0 is true.

   That implies $t_0 \geq t_{n-1,\alpha}$ but $\mu = \mu_0$. Hence the probability of a type I error is

   $$\alpha = \Pr(\text{reject } H_0 | \ H_0 \text{ is true.}) = \Pr(t_0 \geq t_{n-1,\alpha} | \ \mu = \mu_0)$$

   called the *level of significance* or *tolerance limit.*

   The p-value is the type I error when the rejection region on $\bar{X}$ begins exactly its observed value $\bar{x}$.

2. *Type II* error: accept H_0, but in fact H_0 is false.

   With the true value $\mu = \mu_1$, the probability of a type II error is

   $$\begin{aligned} \beta_\alpha(\mu_1) &= \Pr(\text{accept } H_0 | \ H_0 \text{ is false and } \mu = \mu_1) \\ &= \Pr(t_0 < t_{n-1,\alpha} | \ \mu = \mu_1 \geq \mu_0) \end{aligned}$$

   where $\Pr(.| \ \mu_1)$ denotes that this probability is calculated based on the assumption that $\mu = \mu_1$.

This can be summarized into the following table:

Decision	**The fact**	
	H_0 is true	H_0 is false
Accept H_0	O.K.	Type II error
Reject H_0	Type I error	O.K., Power

Type I error is regarded as *more important* because a wrong decision of rejecting H_0 may be fatal. The decision rule is to ensure the type I error to be below α.

A good example is the case of a court. Type I error is the error of charging an innocent person guilty which is more serious than the type II error of releasing a guilty person from penalty. This is due to the respect of human right. Hence a discharged person (fails to reject H_0) can't claim that he is innocent (H_0 is true). In fact, it may just due to insufficient evidence to charge him.

Example: Given $X_i \sim \mathcal{N}(\mu, 10^2)$, $H_0 : \mu = 25$ vs $H_1 : \mu < 25$, $n = 8$, $\bar{x} = 16$, RR $= \{\bar{X} \le k = 18\}$ and $\bar{X} \overset{H_0}{\sim} \mathcal{N}(25, 10^2/8 = 12.5)$. Find

(a) the type I error and type II error when $\mu = 12$.

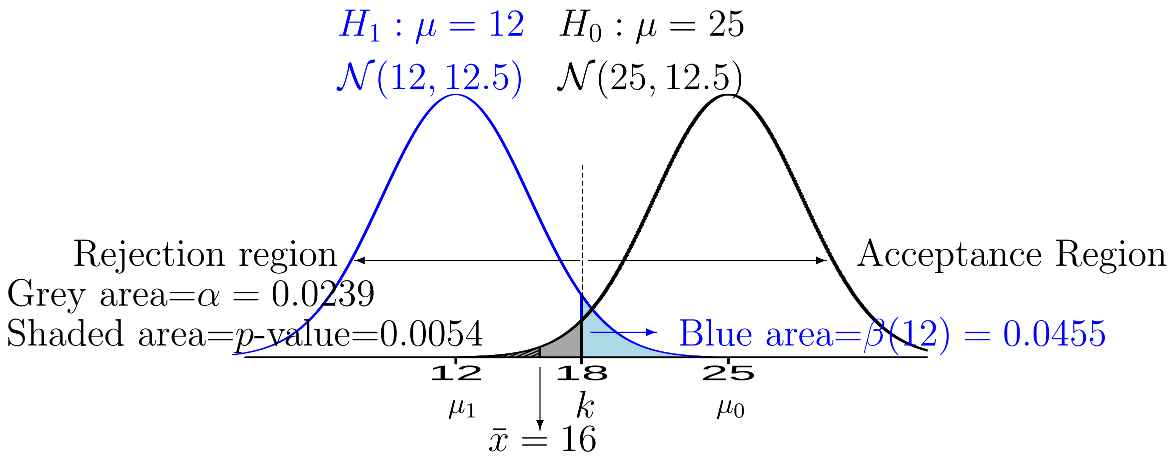
(b) the rejection region and p-value when $\bar{x} = 16$ and $\alpha = 0.05$.

Solution: (a) Under H_0,

$$\begin{aligned} \alpha &= \Pr(\text{type I error}) \\ &= \Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) \\ &= \Pr(\bar{X} \le 18 \mid \bar{X} \sim \mathcal{N}(25, 12.5)) \\ &= \Pr\left(Z < \frac{18-25}{\sqrt{12.5}}\right) = \text{P}(Z < -1.98) \\ &= 1 - 0.9761 = 0.0239. \end{aligned}$$

When $\mu = 12 \in H_1$

$$\begin{aligned} \beta(12) &= \Pr(\text{type II error}) \\ &= \Pr(\text{Accept } H_0 \mid H_0 \text{ is false and } \mu = 12) \\ &= \Pr(\bar{X} > 18 \mid \bar{X} \sim \mathcal{N}(12, 12.5)) \\ &= \Pr\left(Z > \frac{18-12}{\sqrt{12.5}}\right) = \text{P}(Z > 1.69) \\ &= 1 - 0.9545 = 0.0455. \end{aligned}$$



In summary: Given $\bar{x} = 16$ and change the RR determined by k,

k	**Rejection Region**	α	$\beta(12)$	**P-value**	**Conclusion**
19.184	RR=$\{\bar{X} \leq 19.184\}$	0.05	0.0212*	0.0054	Reject H_0
18	RR=$\{\bar{X} \leq 18\}$	0.0239	0.0455	0.0054	Reject H_0
15	RR=$\{\bar{X} \leq 15\}$	0.0023	0.1977	0.0054	Accept H_0

There is a trade-off between α and β. Diminishing the rejection region (RR) will decrease α and increase β and vice versa.

As it is not possible to eliminate both types of errors, we control α, the probability of type I error to be small, say set $\alpha = 0.1, 0.05$, or 0.01 and then achieve a smaller type II errors.

(b) We set $\alpha = 0.05$ and find k for the $RR = \{\bar{X} < k\}$. We know $z_{0.05} = 1.645$. Hence

$$\begin{aligned}\alpha &= \Pr\left(\bar{X} \leq k \mid H_0 \text{ is true }\right) \\ &= \Pr(\bar{X} \leq k \mid \bar{X} \sim \mathcal{N}(25, 12.5)) \\ &= \Pr\left(Z < \frac{k-25}{\sqrt{12.5}}\right).\end{aligned}$$

We know that

$$\frac{k-25}{\sqrt{12.5}} = -1.645 \Rightarrow k = 25 - 1.645 \times \sqrt{12.5} = 19.184.$$

Hence the RR is $\{\bar{X} \leq 19.184\}$ to ensure that $\alpha = 0.05$. Since the sample mean is $\overline{x} = 16$ which lies in RR, we reject H_0. We define

$$\begin{aligned}p\text{-value} &= \Pr\left[\text{Just rej. } H_0 \text{ at } \bar{X} = 16 \text{ or more extreme } |H_0 \text{ is true}\right] \\ &= \Pr[\bar{X} \leq 16 \mid \bar{X} \sim \mathcal{N}(25, 12.5)] \\ &= \Pr\left(Z < \frac{16-25}{\sqrt{12.5}}\right) = \text{P}(Z < -2.55) \\ &= 1 - 0.9946 = 0.0054.\end{aligned}$$

5.2 Power curve of z-test (P.369-374,418-419)

Power of the test at the significance level α for $\mu_1 \in H_1 : \mu > \mu_0$ is

$$\begin{aligned}\text{Power}(\mu_1) &= \Pr(\text{reject } H_0 |\ H_0 \text{ is false and } \mu = \mu_1) \\ &= 1 - \Pr(\text{accept } H_0 |\ \mu = \mu_1) = 1 - \beta_\alpha(\mu_1) = \Pr(z_0 \geq z_\alpha |\ \mu_1).\end{aligned}$$

It is a function of the parameter $\mu_1 \in H_1$, i.e. $\mu_1 > \mu_0$, which gives the probability of rejecting H_0 with the true value $\mu = \mu_1$.

Remarks: Recall that we may have different test statistics for the same test of H_0 vs H_1.

1. With a given significance level α, ideally we would like to have a test such that the power Power(μ_1) is 1 for all $\mu_1 \in H_1$. However it is NOT possible.
2. Instead, we choose the test so that the power is as large as possible for all, or at least for some $\mu_1 \in H_1$. Or we just choose the tests which seem to produce high power.

Example: Suppose that $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, and we wish to test

$$H_0 : \ \mu = \mu_0 \quad \text{vs} \quad H_1 : \ \mu > \mu_0,$$

at the level α. Find the Power of the z-test.

Solution: We have, in general,

$$X_i \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \bar{X} \sim \mathcal{N}(\mu, \sigma^2/n).$$

The critical value at the level α is $z_{1-\alpha}$. Hence the rejection region is

$$\bar{X} \geq \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}.$$

The power of the z-test for any $\mu \in H_1$ is

$$\begin{aligned} \text{Power}(\mu) &= \Pr(\text{reject } H_0 \,|\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)) \\ &= \Pr(\bar{X} \geq \mu_0 + z_{1-\alpha}\sigma/\sqrt{n} \,|\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)) \\ &= \Pr\left(Z \geq \frac{\mu_0 + z_{1-\alpha}\sigma/\sqrt{n} - \mu}{\sigma/\sqrt{n}}\right) \\ &= \Pr\left(Z \geq z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right), \quad \text{for } \mu > \mu_0 = 0. \end{aligned}$$

which is a function of μ given n. In summary, the power curve is

$$\text{For } H_1 : \mu < \mu_0 : \quad \text{Power}(\mu) = \Phi\left(z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right),$$

$$\text{For } H_1 : \mu > \mu_0 : \quad \text{Power}(\mu) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right).$$

For $H_1 : \mu \neq \mu_0$, the power curve is to combine the first curve on the left with the second curve on the right.

When $\mu_0 = 0$, $\sigma = 1$ and $\alpha = 0.05$ such that $z_{0.95} = 1.645$, the power function is

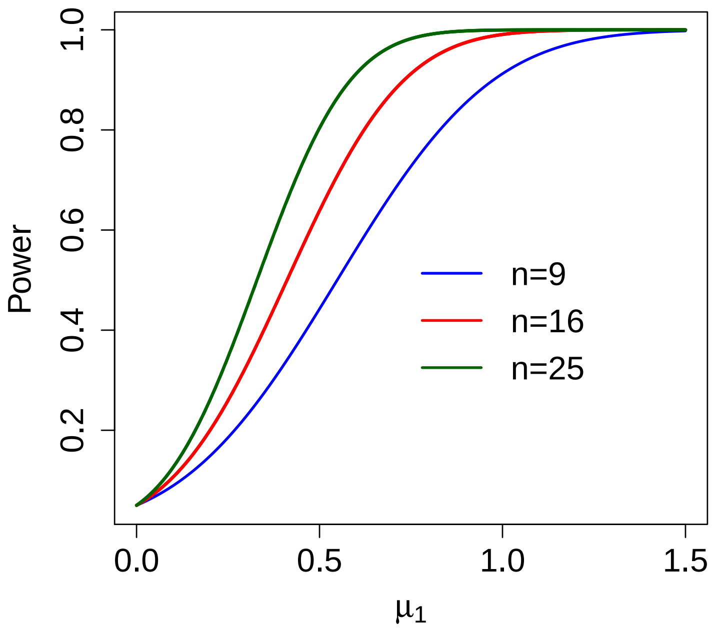
$$\text{Power}(\mu) = 1 - \Phi(1.645 - \sqrt{n}\mu), \quad \text{for } \mu > \mu_0 = 0.$$

Given $n = 9$: Power(0.2)=0.148, Power(0.5)= 0.442, Power(1)=0.912, ...

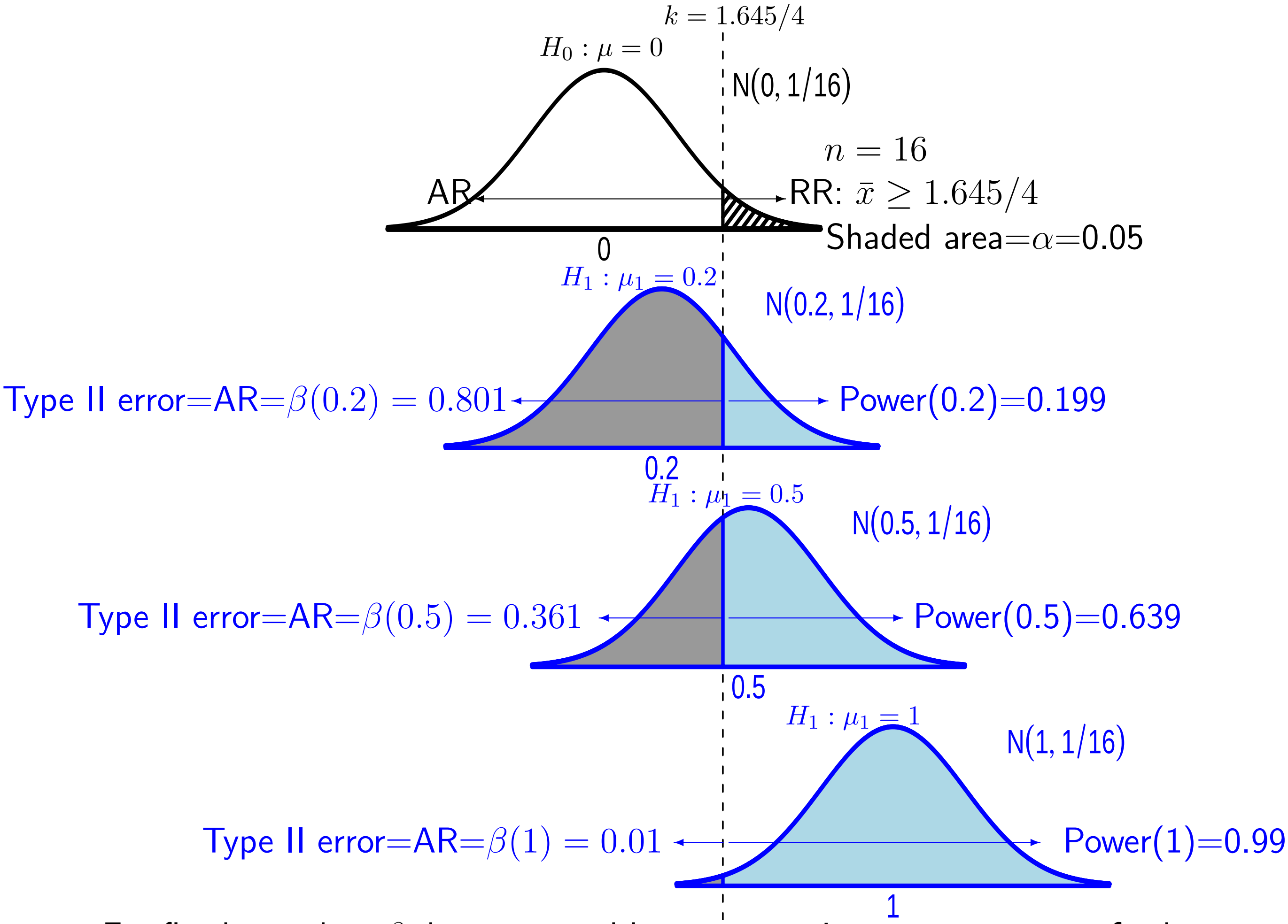
Given $n = 16$: Power(0.2)=0.199, Power(0.5)= 0.639, Power(1)=0.99, ...

Given $n = 25$: Power(0.2)=0.259, Power(0.5)= 0.803, Power(1)=0.999, ...

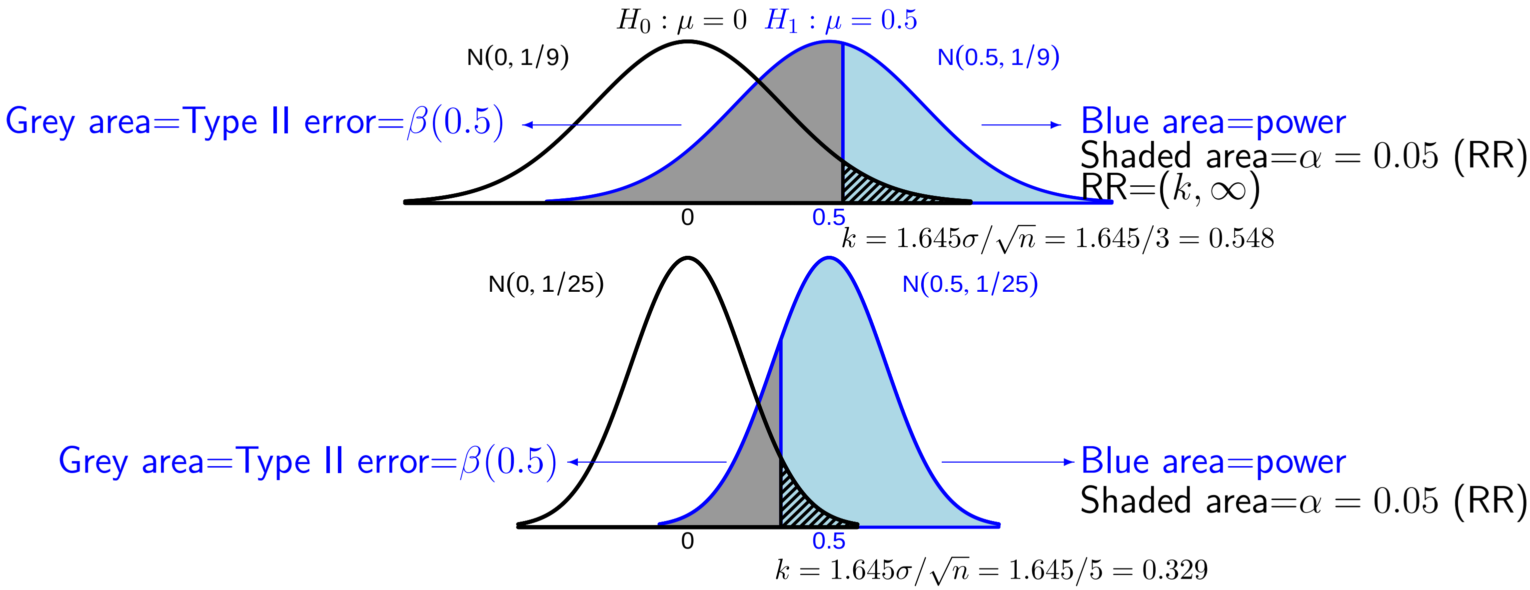
Note: when $n = 9$,
$\text{Power}(0.2) = 1 - \Phi(1.645 - \sqrt{9} \cdot 0.2) = 1 - \Phi(1.045) = 1 - 0.852 = 0.148.$



Power increases with n for fixed μ_1 and increase with $\mu_1 \in H_1 : \mu > 0$ for fixed n.



For fixed n and α, β decreases and hence power increases as μ gets further away from μ_0.



For fixed n and α, β decreases and power increases with increasing n or decreasing σ^2.

Note that the power curve

$$\text{Power}(\mu) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right), \quad \text{if } H_1 : \mu > \mu_0$$

depends on α, $\mu - \mu_0$, n and σ.

There are four ways to increase Power (decrease k where $\text{Power}(\mu) = 1 - \Phi(k)$):

1. **Increase the type I error α**

   $$\nearrow \alpha \quad \searrow z_{1-\alpha} \quad z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \searrow \quad \Phi(\cdot) \searrow \quad 1 - \Phi(\cdot) \nearrow$$

   The higher the tolerance level of type I error α, the easier (higher power) of rejecting H_0 because the evidence required for rejection is less.

2. **Increase μ to get further away from μ_0.**

   $$\nearrow \mu \quad \nearrow \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \quad z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \searrow \quad \Phi(\cdot) \searrow \quad 1 - \Phi(\cdot) \nearrow$$

   The further away is the true mean μ from the hypothesized mean μ_0, the easier (higher power) to detect a difference between them and reject H_0.

3. **Increase the sample size n (more information).**

   $$\nearrow n \quad \searrow \sigma/\sqrt{n} \quad \nearrow \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \quad z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \searrow \quad \Phi(\cdot) \searrow \quad 1 - \Phi(\cdot) \nearrow$$

   The higher the sample size, the more information the sample mean $\bar{x}$ contains to detect a difference from μ_0.

4. **Decrease the population variability σ (more information).**

$$\searrow \sigma \quad \searrow \sigma/\sqrt{n} \quad \nearrow \frac{\mu-\mu_0}{\sigma/\sqrt{n}} \quad z_{1-\alpha}-\frac{\mu-\mu_0}{\sigma/\sqrt{n}} \searrow \quad \Phi(\cdot) \searrow \quad 1-\Phi(\cdot) \nearrow$$

This has the same effect as increasing the sample size: the less variability in the data, the more information it contains about the true mean μ.

5.3 Choice of sample size (P.455-456)

Let us consider the following hypotheses:

$$H_0: \ \mu = \mu_0 \quad \text{vs} \quad H_1: \ \mu > \mu_0.$$

We may be interested in:

> what is the smallest sample size n such that the test at the level α has power at least $100(1-\beta)\%$ when $\mu = \mu_1$?

Example: Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, 1)$, and we wish to test

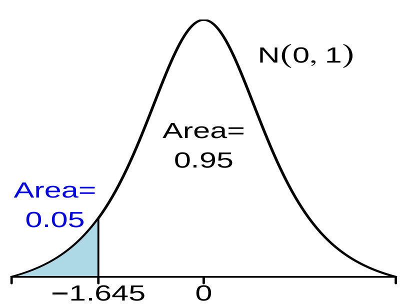
$$H_0: \ \mu = 0 \quad \text{vs} \quad H_1: \ \mu > 0.$$

How large must n be to be at least 99% sure (Power(μ)=0.99) of finding evidence (rej. H_0) at the 5% level ($\alpha = 0.05$) when $\mu = 1$ (H_0 is false)?

Solution: Now $\sigma = 1$, $\mu_0 = 0$ and $z_{0.95} = 1.645$. Using results of the previous example, the Power when $\mu = 1$ for the z-test at the 5% level is

$$\begin{aligned}
& \text{Power}(1) = 1 - \Phi(z_{1-\alpha} - \tfrac{\mu-\mu_0}{\sigma/\sqrt{n}}) && \geq 0.99 \\
\Rightarrow \quad & 1 - \Phi(1.645 - \tfrac{1-0}{1/\sqrt{n}}) && \geq 0.99 \\
\Rightarrow \quad & 1 - \Phi(1.645 - \sqrt{n} \cdot 1) && \geq 0.99 \\
\Rightarrow \quad & \Phi(1.645 - \sqrt{n}) && \leq 0.01 \\
\Rightarrow \quad & 1.645 - \sqrt{n} && \leq -2.326 \\
\Rightarrow \quad & \sqrt{n} && \geq 1.645 + 2.326 \\
\Rightarrow \quad & n && \geq (3.971)^2 = 15.769.
\end{aligned}$$

So we need $n = 16$.



6 Confidence intervals

6.1 Definition

The *point* estimate $\hat{\theta}$ (say $\bar{x}$) of a parameter θ (say μ) does not show its variability across samples. To show such estimation precision, we should find an *interval* estimate.

Definition: Let $\hat{\theta}_L$ and $\hat{\theta}_R$ be two statistics. If

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_R) = 1 - \alpha,$$

then the random interval $[\hat{\theta}_L, \hat{\theta}_R]$ is called a $100(1-\alpha)\%$ *confidence interval* (CI) for θ, and $100(1-\alpha)\%$ is called the *confidence level* of the interval.

Depending on whether the end points are included, the CI could be

$$[\hat{\theta}_L, \hat{\theta}_R), \quad (\hat{\theta}_L, \hat{\theta}_R], \quad (\hat{\theta}_L, \hat{\theta}_R)$$

which exclude θ_R, θ_L and both respectively.

In general, the α may be chosen to be $0.01, 0.05, 0.10$, etc, and then we get $99\%, 95\%, 90\%$ confidence interval accordingly.

Remarks:

1. Depending on test statistic (say sample mean or median), there are many CIs for θ. With greater estimation precision, the CI should be narrower.

2. The *one-sided* CIs are $(-\infty, \hat{\theta}_R)$ and $(\hat{\theta}_L, \infty)$ where the $\hat{\theta}_R$ and $\hat{\theta}_L$ are the $100(1-\alpha)\%$ *upper* bound for θ, and $100(1-\alpha)\%$ *lower* bound for θ, respectively.

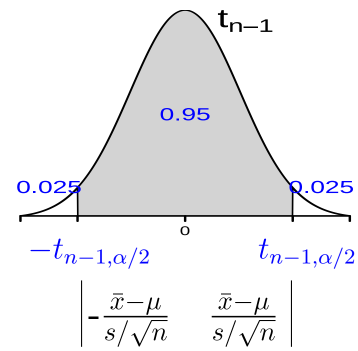
6.2 Confidence intervals for the mean (P.396-400)

Let $X_1, X_2, ..., X_n$ be a random sample from normal population and $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown. Then $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ and

$$\Pr\left[-t_{n-1,\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1,\alpha/2}\right] = 1-\alpha$$

$$\Rightarrow \quad \Pr\left[-t_{n-1,\alpha/2} < \frac{\mu-\bar{X}}{S/\sqrt{n}} < t_{n-1,\alpha/2}\right] = 1-\alpha$$

$$\Rightarrow \quad \Pr[\bar{X} - t_{n-1,\alpha/2}S/\sqrt{n} < \mu < \bar{X} + t_{n-1,\alpha/2}S/\sqrt{n}] = 1-\alpha$$



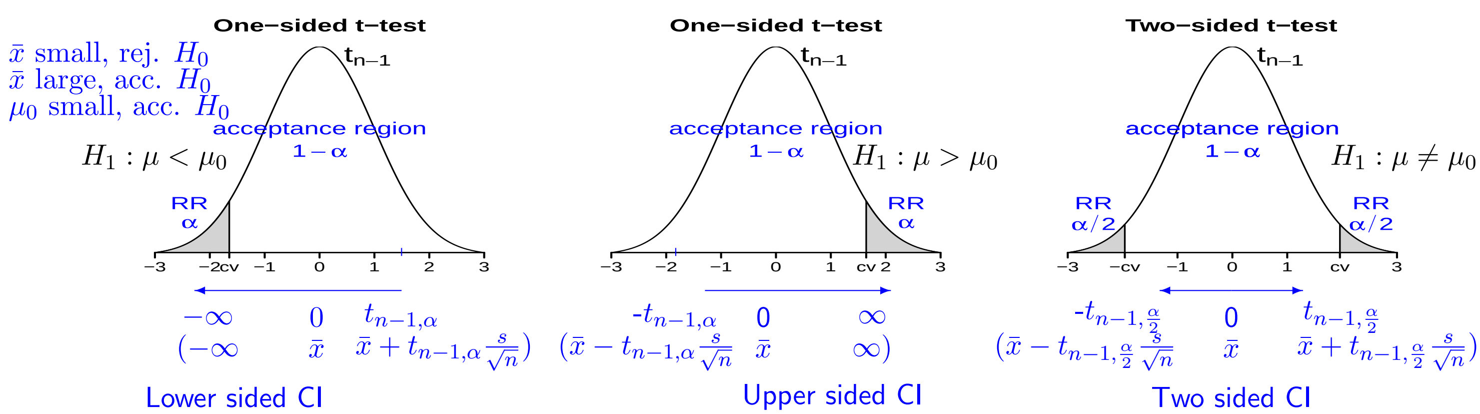
Similarly, for the test with $H_1 : \mu > \mu_0$, we accept H_0 if

$$t_0 < t_{n-1}(\alpha) \;\Rightarrow\; \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < t_{n-1,\alpha} \quad \Rightarrow\; \frac{\mu_0-\bar{x}}{s/\sqrt{n}} > -t_{n-1,\alpha}$$

$$\Rightarrow \mu_0 > \bar{x} - t_{n-1,\alpha}\tfrac{s}{\sqrt{n}} \;\Rightarrow\; \mu_0 \in (\bar{x} - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \infty).$$

Hence a $100(1-\alpha)\%$ confidence interval for the mean μ is

$$\begin{array}{ll} \text{Lower sided CI } H_1 : \mu < \mu_0 : & (-\infty,\; \bar{x} + t_{\alpha,n-1}s/\sqrt{n}), \\ \text{Upper sided CI } H_1 : \mu > \mu_0 : & (\bar{x} - t_{\alpha,n-1}s/\sqrt{n},\; \infty), \\ \text{Two sided CI } H_1 : \mu \neq \mu_0 : & (\bar{x}-t_{\alpha/2,n-1}s/\sqrt{n},\; \bar{x}+t_{\alpha/2,n-1}s/\sqrt{n}) \end{array}$$



When σ^2 is know, the two-sided and one-sided CIs are given by:

$$\begin{array}{ll} \text{Lower sided CI } H_1: \mu < \mu_0: & (-\infty,\ \bar{x} + z_\alpha \sigma/\sqrt{n}), \\ \text{Upper sided CI } H_1: \mu > \mu_0: & (\bar{x} - z_\alpha \sigma/\sqrt{n},\ \infty), \\ \text{Two sided CI } H_1: \mu \neq \mu_0: & (\bar{x} - z_{\alpha/2} \sigma/\sqrt{n},\ \bar{x} + z_{\alpha/2} \sigma/\sqrt{n}) \end{array}$$

6.3 Meaning of confidence interval

Suppose a 95% confidence interval for the mean μ is (a, b). This does *not* mean that

1. 95% of the means μ are in (a, b), that is $\Pr(a < \mu < b) = 0.95$ since μ is a *fixed* but unknown parameter nor
2. $\Pr(a < \bar{X} < b) = 0.95$, where $\bar{X}$ is the sample mean since the CI is for the true mean μ not the sample mean $\bar{X}$.

It means that if we draw a large number of random samples and compute for each sample a 95% CI, about 95% of these CIs will contain μ. The following example demonstrates the meaning of CI.

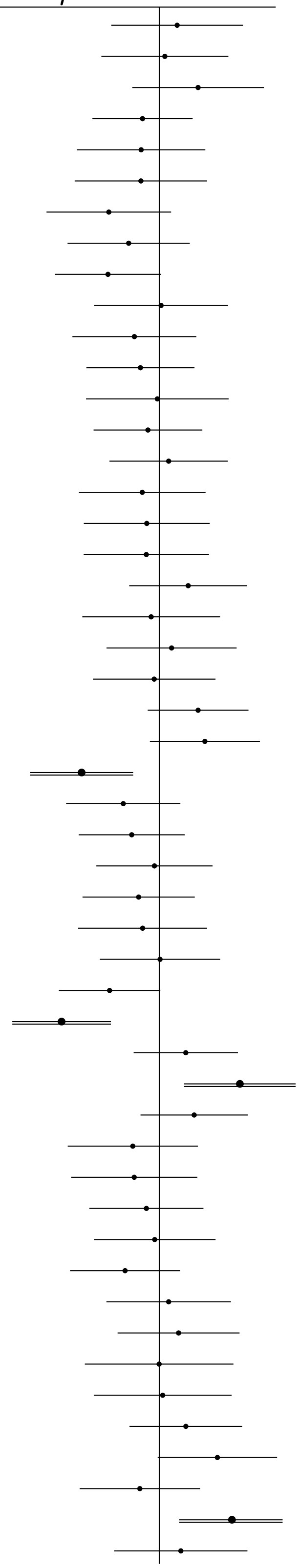
Example: The following is a set of $N = 100$ values of a population with $\mu = 52.575$ and $\sigma^2 = 886.847$.

67.8	47.7	89.1	74.8	90.0	7.1	91.8	66.0	39.1	30.0	26.7	21.7	93.7	62.0	70.2
5.4	3.7	68.4	19.6	50.8	91.8	27.2	67.2	60.1	73.1	2.0	30.1	88.2	84.0	81.3
85.9	35.5	24.3	62.7	30.0	97.8	42.0	51.7	20.4	48.4	13.7	33.0	91.7	57.9	19.9
1.8	19.8	84.7	63.7	83.8	87.4	21.4	83.0	60.9	23.9	8.5	50.7	35.7	92.4	89.4
64.8	11.5	67.2	6.0	21.1	11.7	12.7	81.9	4.6	22.4	87.3	72.6	86.9	73.2	44.2
18.5	21.4	84.2	98.0	63.5	25.3	75.2	22.9	29.3	83.3	85.1	58.9	80.0	93.7	29.0
83.2	76.7	73.9	98.1	23.9	32.7	29.9	23.3	45.6	82.5					

Random samples of size $n = 20$ are drawn from these 100 values repeatedly 50 times. The 50 CIs and their coverages of the true mean μ are then shown.

$\bar{x}$	s^2	C.I.
56.020	1047.629	(43.332, 68.708)
53.650	973.679	(41.418, 65.882)
60.052	1044.769	(47.381, 72.722)
49.350	606.324	(39.697, 59.002)
49.082	994.433	(36.721, 61.444)
49.038	1058.878	(36.282, 61.794)
42.857	937.009	(30.858, 54.856)
46.682	901.619	(34.911, 58.453)
42.694	677.978	(32.487, 52.901)
52.922	1086.781	(39.999, 65.844)
47.778	926.727	(35.845, 59.712)
48.950	705.443	(38.539, 59.362)
52.200	1227.258	(38.467, 65.933)
50.395	714.205	(39.919, 60.871)
54.384	845.914	(42.982, 65.785)
49.296	968.221	(37.099, 61.494)
50.167	957.080	(38.040, 62.295)
50.082	948.243	(38.010, 62.153)
58.146	840.061	(46.785, 69.508)
51.010	1144.449	(37.749, 64.271)
54.947	1021.469	(42.418, 67.476)
51.596	907.564	(39.787, 63.405)
60.053	612.693	(50.350, 69.756)
61.360	730.304	(50.767, 71.954)
37.612	642.730	(27.674, 47.550)
45.641	788.646	(34.632, 56.640)
47.266	678.076	(37.059, 57.474)
51.645	815.394	(40.452, 62.839)
48.601	760.584	(37.790, 59.412)
49.368	1003.110	(36.953, 61.784)
52.723	874.174	(41.133, 64.313)
43.005	622.081	(33.228, 52.782)
33.760	586.996	(24.262, 43.257)
57.683	656.446	(47.639, 67.726)
68.100	750.229	(57.363, 78.837)
59.298	695.199	(48.962, 69.634)
47.474	1021.986	(34.942, 60.006)
47.749	962.295	(35.588, 59.909)
50.098	785.590	(39.111, 61.085)
51.697	893.741	(39.978, 63.416)
45.989	731.062	(35.390, 56.588)
54.382	735.614	(42.392, 66.373)
56.294	898.002	(44.547, 68.041)
52.548	1333.015	(38.236, 66.860)
53.236	1147.398	(39.958, 66.514)
57.694	766.730	(46.840, 68.548)
63.771	860.750	(52.270, 75.271)
48.835	875.848	(37.234, 60.437)
66.575	645.377	(56.416, 76.333)
56.731	1070.385	(43.906, 69.556)

$\mu = 52.575$



46 CIs cover μ

4 CIs not cover μ

2 on left & 2 on right

Actual coverage

$= \dfrac{46}{50} = 0.92$

Nominal coverage

$= 1 - \alpha = 0.95$

Note that the CIs change in both location and length as we move from sample to sample. Hence CI is also *random* and in repeated sampling, roughly 95% of the intervals contain the true mean μ.

6.4 Relationships between acceptance region and confidence interval

For testing the mean μ with 2-sided $H_1: \ \mu \neq \mu_0$ at level α, we use the t-test with test statistic

$$t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

and reject $H_0: \ \mu = \mu_0$ if t_0 falls in the RR $\{t_0: \ |t_0| \geq t_{n-1}(\alpha/2)\}$. Then the *acceptance region (AR)* for the *sample mean* $\bar{x}$ is

$$\begin{aligned} |t_0| &= \left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| < t_{n-1,\alpha/2} \\ \Leftrightarrow \ \bar{x} &\in (\ \mu_0 - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \ \mu_0 + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\ ). \end{aligned}$$

Hence μ_0 is the center and $\bar{x}$ is used to test.

Alternatively, the *region (CI)* for the *true mean* μ is

$$\begin{aligned} |t_0| &= \left|\frac{\mu_0 - \bar{x}}{s/\sqrt{n}}\right| < t_{n-1,\alpha/2} \\ \Leftrightarrow \ \mu_0 &\in (\ \bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \ \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\ ) = \text{CI for } \mu. \end{aligned}$$

Hence $\bar{x}$ is the center and μ_0 is used to test.

Thus, the decision rule using CI is

> accept H_0 if μ_0 lies in the $100(1-\alpha)\%$ CI for the μ and reject H_0 otherwise.

Hence the *decision rule* using a $100(1-\alpha)\%$ CI for the μ is

$$\text{accept } H_0 \text{ if } \mu_0 \in \left(\bar{x} - t_{n-1,\alpha}\frac{s}{\sqrt{n}},\ \infty\right) \quad \text{if } H_1: \ \mu > \mu_0;$$

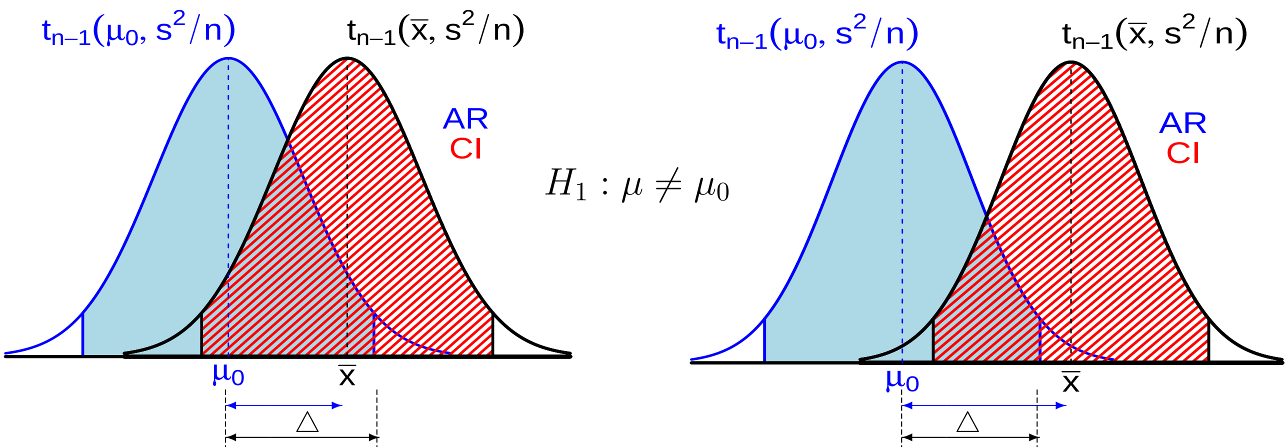
$$\text{accept } H_0 \text{ if } \mu_0 \in \left(-\infty,\ \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}}\right) \quad \text{if } H_1: \ \mu < \mu_0;$$

$$\text{accept } H_0 \text{ if } \mu_0 \in \left(\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right) \quad \text{if } H_1: \ \mu \neq \mu_0;$$

reject H_0, if μ_0 lies outside the CI.

Equivalently, a $100(1-\alpha)\%$ CI for μ can be interpreted as the set of all values of μ_0 for which $H_0: \ \mu = \mu_0$ is *acceptable* at level α.

When σ^2 is known, we perform the z-test. The *decision rule* will follow those above with z_b replacing $t_{n-1,b}$ and σ replacing s where $b = \alpha$ for one-sided test and $b = \alpha/2$ for two-sided test.



If $|\mu_0 - \bar{x}| < t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} = \triangle$, accept H_0

$\bar{x} \in (\mu_0 - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \mu_0 + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})$=AR

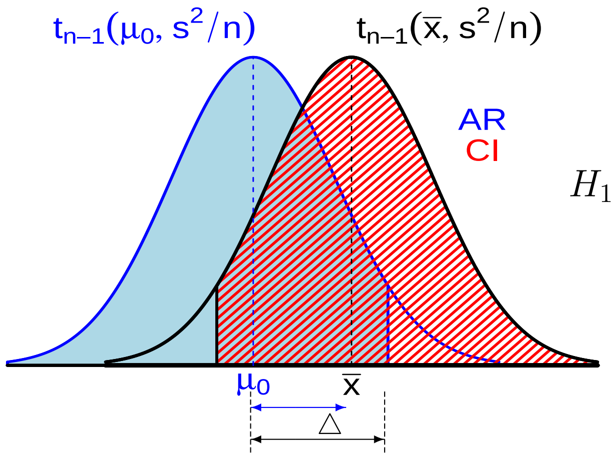
$\mu_0 \in (\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})$ = CI

If $|\mu_0 - \bar{x}| > t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$, reject H_0

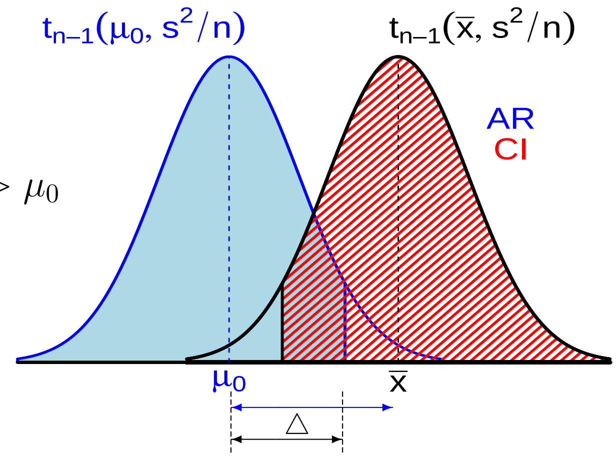
$\bar{x} \notin (\mu_0 - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \mu_0 + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})$=AR

$\mu_0 \notin (\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})$ = CI

$H_1 : \mu > \mu_0$

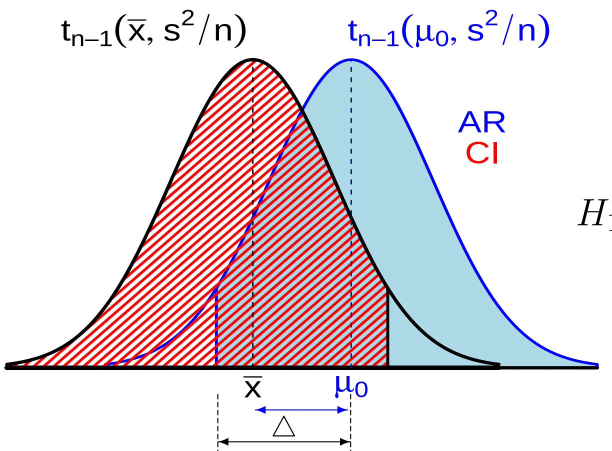


If $|\mu_0 - \bar{x}| < t_{n-1,\alpha}\frac{s}{\sqrt{n}} = \triangle$, accept H_0
$\bar{x} \in (-\infty, \mu_0 + t_{n-1,\alpha}\frac{s}{\sqrt{n}})$ = AR
$\mu_0 \in (\bar{x} - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \infty)$ = CI

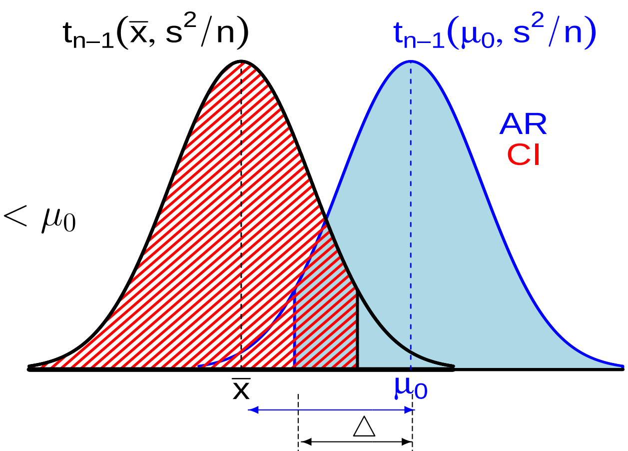


If $|\mu_0 - \bar{x}| > t_{n-1,\alpha}\frac{s}{\sqrt{n}}$, reject H_0
$\bar{x} \notin (-\infty, \mu_0 + t_{n-1,\alpha}\frac{s}{\sqrt{n}})$ = AR
$\mu_0 \notin (\bar{x} - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \infty)$ = CI

$H_1 : \mu < \mu_0$



If $|\mu_0 - \bar{x}| < t_{n-1,\alpha}\frac{s}{\sqrt{n}} = \triangle$, accept H_0
$\bar{x} \in (\mu_0 - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \infty)$ = AR
$\mu_0 \in (-\infty, \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}})$ = CI



If $|\mu_0 - \bar{x}| > t_{n-1,\alpha}\frac{s}{\sqrt{n}}$, reject H_0
$\bar{x} \notin (\mu_0 - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \infty)$ = AR
$\mu_0 \notin (-\infty, \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}})$ = CI

Example: (Beer contents) A brand of beer claims its beer content is 375 (in millilitres) on the label. A sample of 40 bottles of the beer gave a sample average of 373.9 and a standard deviation of 2.5. Is there evidence that the mean content of the beer is less than the 375 mL as claimed on the label at level $\alpha = 0.05$?

(a) Construct a 95% two-sided confidence interval for the mean and a 95% one-sided confidence interval for the mean for testing the claim.

(b) Test the claim on the label using the CI in (a).

Solution: We have $n = 40$, $\bar{x} = 373.9$, $s = 2.5$, $\mu_0 = 375$, $H_1 : \mu < 375$, $t_0 = -2.78$, $t_{39,0.05} = 1.684$ and $t_{39,0.025} = 2.021$.

(a) The 95% two-sided CI for the true average beer content μ is

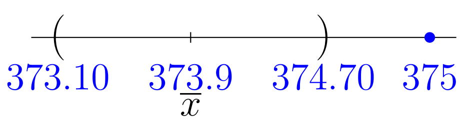
$$\begin{aligned} & (\bar{x} - t_{n-1,\alpha/2} s/\sqrt{n}, \bar{x} + t_{n-1,\alpha/2} s/\sqrt{n}) \\ = & (373.9 - 2.021 \cdot 2.5/\sqrt{40},\ 373.9 + 2.021 \cdot 2.5/\sqrt{40}) \\ = & (373.1011,\ 374.6989). \end{aligned}$$

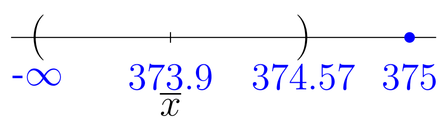
The 95% lower-sided CI for the true average beer content μ is

$$\text{RR: } t_0 < -t_{n-1,\alpha} \Leftrightarrow \text{AR: } t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \in (-t_{n-1,\alpha},\ \infty)$$

$$\Leftrightarrow \text{CI: } \frac{\mu_0 - \bar{x}}{s/\sqrt{n}} \in (-\infty,\ t_{n-1,\alpha}) \Leftrightarrow \mu_0 \in (-\infty,\ \bar{x} + t_{n-1,\alpha}\, s/\sqrt{n})$$

$$\begin{aligned} (-\infty,\ \bar{x} + t_{n-1,\alpha}\, s/\sqrt{n}) &= (-\infty,\ 373.9 + 1.684 \cdot 2.5/\sqrt{40}) \\ &= (-\infty,\ 374.57). \end{aligned}$$



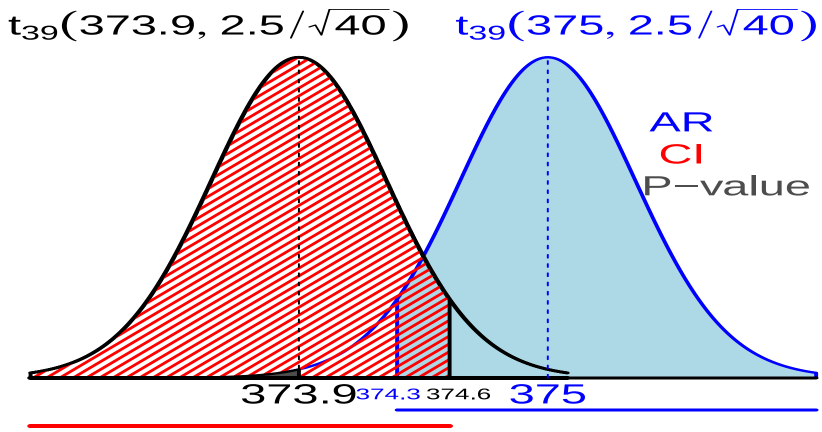


(b) The one-sample t-test for the mean beer content μ is

1. **Hypothesis:** $H_0 : \mu = 375$ vs $H_1 : \mu < 375$.

4''. **CI for accepting** H_0: $\mu_0 \in (-\infty, 374.57)$.

5. **Decision**: Since $\mu_0 = 375 \notin (-\infty, 374.57)$, there is strong evidence in the data against H_0. The mean content of the beer is less than the 375 mL at the level $\alpha = 0.05$.



On standardized scale: $t_0 = -2.78 < -1.68$. In RR. Reject H_0.

On Measurement scale: $\bar{x} = 373.9 \notin (374.3, \infty) =$ AR. Reject H_0.

On Measurement scale: $\mu_0 = 375 \notin (-\infty, 374.6) =$ CI. Reject H_0.

1-sided area $=$ p-value=0.0041<0.05. Reject H_0.

Note:

$\Delta = t_{n-1,\alpha}\frac{s}{\sqrt{n}} = 1.684 \times \frac{2.5}{\sqrt{40}} = 0.666$.

$\mu_0 - \bar{x} = 375 - 373.9 = 1.1 > 0.666$. Sufficient evidence against H_0.