Part II

NORMAL DISTRIBUTION

In this part we cover classical tests such as the t-test for the mean and the χ^2-test for the population variance. We assume throughout this part that the underlying distribution is Gaussian. Chapter 2 covers tests for the questions if a mean equals a specific value or if two populations share the same mean. Chapter 3 presents statistical tests on variances of one or two normal populations. In both chapters it must be ascertained whether the accompanying parameters (the variance for the mean tests and the mean for the variance tests) are known or unknown. In the two sample cases it is also necessary to determine whether the two samples are independent or not.

2

Tests on the mean

This chapter contains statistical tests on the mean of a normal population. Frequent questions are if the mean equals a specific value (mostly the null) or if two populations have the same mean or differ by a specific value. Depending on the sampling strategy and on knowledge of the data generation process, the assumptions of known or unknown variances must be distinguished. In most situations the variance is unknown–probably the reason why neither SAS nor R provides procedures to calculate tests for the case of known variances. However, rare situations exist where the variance of the underlying Gaussian distribution is known. We provide some code that demonstrates how these–so-called z-tests–can be calculated. If the variance has to be estimated from the sample, the test statistic distribution changes from standard normal to Student's t-distribution. Here the degrees of freedom vary depending on the specific test problem, for example, in the two population case on whether the variances are assumed to be equal or not. SAS by the procedure `proc ttest` and R by the function `t.test` provided convenient ways to calculate these tests. For k-sample tests (F-test) please refer to Chapter 17 which covers ANOVA tests.

2.1   One-sample tests

In this section we deal with the question, if the mean of a normal population differs from a predefined value. Whether the variance of the underlying Gaussian distribution is known or not determines the use of the z-test or the t-test.

2.1.1   z-test

Description: Tests if a population mean μ differs from a specific value μ_0.

Assumptions: • Data are measured on an interval or ratio scale.

• Data are randomly sampled from a Gaussian distribution.

• Standard deviation σ of the underlying Gaussian distribution is known.

Hypotheses: (A) $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
(B) $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$
(C) $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$

Test statistic: $Z = \frac{\overline{X} - \mu_0}{\sigma} \sqrt{n}$ with $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Test decision: Reject H_0 if for the observed value z of Z
(A) $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$
(B) $z > z_{1-\alpha}$
(C) $z < z_{\alpha}$

p-value: (A) $p = 2\Phi(-|z|)$
(B) $p = 1 - \Phi(z)$
(C) $p = \Phi(z)$

Annotations:
- The test statistic Z follows a standard normal distribution.

- z_{α} is the α-quantile of the standard normal distribution.

- The assumption of an underlying Gaussian distribution can be relaxed if the sample size is large. Usually a sample size $n \geq 25$ or 30 is considered to be large enough.

Example: To test the hypothesis that the mean systolic blood pressure in a certain population equals 140 mmHg. The standard deviation has a known value of 20 and a data set of 55 patients is available (dataset in Table A.1).

SAS code

```
data blood_pressure;
 set c.blood_pressure;
run;

* Calculate sample mean and total sample size;
proc means data=blood_pressure mean std;
 var mmhg;
 output out=ztest01 mean=meanvalue n=n_total;
run;

* Calculate test-statistic and p-values;
data ztest02;
 set ztest01;
 format p_value_A p_value_B p_value_C pvalue.;
 mu0=140;     * Set mean value under the null hypothesis;
 sigma=20;    * Set known sigma;
 z=sqrt(n_total)*(meanvalue-mu0)/sigma;

 p_value_A=2*probnorm(-abs(z));
 p_value_B=1-probnorm(z);
 p_value_C=probnorm(z);
run;
```

```
* Output results;
proc print;
 var z p_value_A p_value_B p_value_C ;
run;
```

SAS output

```
     z       p_value_A   p_value_B   p_value_C
 -3.70810    0.0002      0.9999      0.0001
```

Remarks:

- There is no SAS procedure to calculate the one-sample z-test directly.

- The above code also shows how to calculate the p-values for the one-sided tests (B) and (C).

R code

```
# Calculate sample mean and total sample size
xbar<-mean(blood_pressure$mmhg)
n<-length(blood_pressure$mmhg)

# Set mean value under the null hypothesis
mu0<-140

# Set known sigma
sigma<-20

# Calculate test statistic and p-values
z<-sqrt(n)*(xbar-mu0)/sigma

p_value_A=2*pnorm(-abs(z))
p_value_B=1-pnorm(z)
p_value_C=pnorm(z)

# Output results
z
p_value_A
p_value_B
p_value_C
```

R output

```
> z
[1] -3.708099
> p_value_A
[1] 0.0002088208
> p_value_B
[1] 0.9998956
> p_value_C
[1] 0.0001044104
```

Remarks:

- There is no basic R function to calculate the one-sample z-test directly.

- The above code also shows how to calculate the p-values for the one-sided tests (B) and (C).

2.1.2   t-test

Description: Tests if a population mean μ differs from a specific value μ_0.

Assumptions:
- Data are measured on an interval or ratio scale.

- Data are randomly sampled from a Gaussian distribution.

- Standard deviation σ of the underlying Gaussian distribution is unknown and estimated by the population standard deviation s.

Hypotheses: (A) $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
(B) $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$
(C) $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$

Test statistic: $T = \frac{\overline{X} - \mu_0}{s} \sqrt{n}$ with $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$

Test decision: Reject H_0 if for the observed value t of T
(A) $t < t_{\alpha/2, n-1}$ or $t > t_{1-\alpha/2, n-1}$
(B) $t > t_{1-\alpha, n-1}$
(C) $t < t_{\alpha, n-1}$

p-value: (A) $p = 2 \, P(T \leq (-|t|))$
(B) $p = 1 - P(T \leq t))$
(C) $p = P(T \leq t))$

Annotations:
- The test statistic T is t-distributed with $n - 1$ degrees of freedom.

- $t_{\alpha, n-1}$ is the α-quantile of the t-distribution with $n - 1$ degrees of freedom.

- The assumption of an underlying Gaussian distribution can be relaxed if the sample size is large. Usually a sample size $n \geq 30$ is considered to be large enough.

Example: To test the hypothesis that the mean systolic blood pressure in a certain population equals 140 mmHg. The dataset at hand has measurements on 55 patients (dataset in Table A.1).

SAS code

```
proc ttest data=blood_pressure ho=140 sides=2;
 var mmhg;
run;
```

SAS output

```
  DF     t Value    Pr < t
  54       -3.87    0.0003
```

Remarks:

- ho=*value* is optional and defines the value μ_0 to test against. Default is 0.

- sides=*value* is optional and defines the type of alternative hypothesis: 2=two sided (A); U=true mean is greater (B); L=true mean is lower (C). Default is 2.

R code

```
t.test(blood_pressure$mmhg,mu=140,alternative="two.sided")
```

R output

```
t = -3.8693, df = 54, p-value = 0.0002961
```

Remarks:

- mu=*value* is optional and defines the value μ_0 to test against. Default is 0.

- alternative=*"value"* is optional and defines the type of alternative hypothesis: "two.sided"= two sided (A); "greater"=true mean is greater (B); "less"=true mean is lower (C). Default is "two.sided".

2.2   Two-sample tests

This section covers two-sample tests, to test if either the means of two populations differ from each other or if the mean difference of paired populations differ from a specific value.

2.2.1   Two-sample z-test

Description: Tests if two population means μ_1 and μ_2 differ less than, more than or by a value d_0.

Assumptions:
- Data are measured on an interval or ratio scale.
- Data are randomly sampled from two independent Gaussian distributions.
- The standard deviations σ_1 and σ_2 of the underlying Gaussian distributions are known.

Hypotheses:
(A) $H_0 : \mu_1 - \mu_2 = d_0$ vs $H_1 : \mu_1 - \mu_2 \neq d_0$
(B) $H_0 : \mu_1 - \mu_2 \leq d_0$ vs $H_1 : \mu_1 - \mu_2 > d_0$
(C) $H_0 : \mu_1 - \mu_2 \geq d_0$ vs $H_1 : \mu_1 - \mu_2 < d_0$

Test statistic:
$Z = [(\overline{X}_1 - \overline{X}_2) - d_0] \Big/ \sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$

Test decision: Reject H_0 if for the observed value z of Z
(A) $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$
(B) $z > z_{1-\alpha}$
(C) $z < z_\alpha$

p-value:
(A) $p = 2\Phi(-|z|)$
(B) $p = 1 - \Phi(z)$
(C) $p = \Phi(z)$

Annotations:
- The test statistic Z is a standard normal distribution.
- z_α is the α-quantile of the standard normal distribution.
- The assumption of an underlying Gaussian distribution can be relaxed if the sample size is large. Usually sample sizes $n_1, n_2 \geq 25$ or 30 for both distributions are considered to be large enough.

Example: To test the hypothesis that the mean systolic blood pressures of healthy subjects (status=0) and subjects with hypertension (status=1) are equal ($d_0 = 0$) with known standard deviations of $\sigma_1 = 10$ and $\sigma_2 = 12$. The dataset contains $n_1 = 25$ subjects with status 0 and $n_2 = 30$ with status 1 (dataset in Table A.1).

SAS code

```
* Calculate the two means and sample sizes
proc means data=blood_pressure mean;
 var mmhg;
 by status;
 output out=ztest01 mean=meanvalue n=n_total;
run;

* Output of the means in two different datasets;
data ztest02 ztest03;
 set ztest01;
 if status=0 then output ztest02;
```

```
 if status=1 then output ztest03;
run;

* Rename mean and sample size of subjects
                                with status=0;
data ztest02;
 set ztest02;
 rename meanvalue=mean_status0
        n_total=n_status0;
run;

* Rename mean and sample size of subjects
                                with status=1;
data ztest03;
 set ztest03;
 rename meanvalue=mean_status1
        n_total=n_status1;
run;

* Calculate test statistic and two-sided p-value;
data ztest04;
 merge ztest02 ztest03;
 * Set difference to be tested;
 d0=0;
 * Set standard deviation of sample with status 0;
 sigma0=10;
 * Set standard deviation of sample with status 1;
 sigma1=12;
 format p_value pvalue.;
 z= ((mean_status0-mean_status1)-d0)
     / sqrt(sigma0**2/n_status0+sigma1**2/n_status1);
 p_value=2*probnorm(-abs(z));
run;

* Output results;
proc print;
 var z p_value;
run;
```

SAS output

```
    z       p_value
 -10.5557   <.0001
```

Remarks:

- There is no SAS procedure to calculate the two-sample z-test directly.

- The one-sided p-value for hypothesis (B) can be calculated with `p_value_B=1-probnorm(z)` and the p-value for hypothesis (C) with `p_value_C=probnorm(z)`.

R code

```
# Set difference to be tested;
d0<-0
# Set standard deviation of sample with status 0
sigma0<-10
# Set standard deviation of sample with status 1
sigma1<-12
# Calculate the two means
mean_status0<-
   mean(blood_pressure$mmhg[blood_pressure$status==0])
mean_status1<-
   mean(blood_pressure$mmhg[blood_pressure$status==1])
# Calculate both sample sizes
n_status0<-
  length(blood_pressure$mmhg[blood_pressure$status==0])
n_status1<-
  length(blood_pressure$mmhg[blood_pressure$status==1])
# Calculate test statistic and two-sided p-value
z<-((mean_status0-mean_status1)-d0)/
   sqrt(sigma0^2/n_status0+sigma1^2/n_status1)
p_value=2*pnorm(-abs(z))
# Output results
z
p_value
```

R output

```
> z
[1] -10.55572
> p_value
[1] 4.779482e-26
```

Remarks:

- There is no basic R function to calculate the two-sample z-test directly.

- The one-sided p-value for hypothesis (B) can be calculated with `p_value_B=1-pnorm(z)` and the p-value for hypothesis (C) with `p_value_C=pnorm(z)`.

2.2.2   Two-sample pooled t-test

Description: Tests if two population means μ_1 and μ_2 differ less than, more than or by a value d_0.

Assumptions:
- Data are measured on an interval or ratio scale.

- Data are randomly sampled from two independent Gaussian distributions.

- Standard deviations σ_1 and σ_2 of the underlying Gaussian distributions are unknown but equal and estimated through the pooled population standard deviation s_p.

Hypotheses:
(A) $H_0 : \mu_1 - \mu_2 = d_0$ vs $H_1 : \mu_1 - \mu_2 \neq d_0$
(B) $H_0 : \mu_1 - \mu_2 \leq d_0$ vs $H_1 : \mu_1 - \mu_2 > d_0$
(C) $H_0 : \mu_1 - \mu_2 \geq d_0$ vs $H_1 : \mu_1 - \mu_2 < d_0$

Test statistic:
$$T = [(\overline{X}_1 - \overline{X}_2) - d_0] \Big/ \left[s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

$$\text{with } s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}},$$

$$\text{where } s_j = \sqrt{\frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_i - \overline{X}_j)^2}, \text{ for } j = 1, 2.$$

Test decision:
Reject H_0 if for the observed value t of T
(A) $t < t_{\alpha/2, n_1+n_2-2}$ or $t > t_{1-\alpha/2, n_1+n_2-2}$
(B) $t > t_{1-\alpha, n_1+n_2-2}$
(C) $t < t_{\alpha, n_1+n_2-2}$

p-value:
(A) $p = 2\ P(T \leq (-|t|))$
(B) $p = 1 - P(T \leq t))$
(C) $p = P(T \leq t))$

Annotations:
- The test statistic T is t-distributed with $n_1 + n_2 - 2$ degrees of freedom.

- t_{α, n_1+n_2-2} is the α-quantile of the t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

- The assumption of two underlying Gaussian distributions can be relaxed if the sample sizes of both samples are large. Usually sample sizes $n_1, n_2 \geq 25$ or 30 for both distributions are considered to be large enough.

Example: To test the hypothesis that the mean systolic blood pressures of healthy subjects (status=0) and subjects with hypertension (status=1) are equal, hence $d_0 = 0$. The dataset contains $n_1 = 25$ subjects with status 0 and $n_2 = 30$ with status 1 (dataset in Table A.1).

SAS code

```
proc ttest data=blood_pressure h0=0 sides=2;
 class status;
 var mmhg;
run;
```

SAS output

```
 Method  Variances  DF  t Value  Pr > |t|
 Pooled    Equal    53  -10.47   <.0001
```

Remarks:

- h0=*value* is optional and defines the value μ_0 to test against. Default is 0.

- sides=*value* is optional and defines the type of alternative hypothesis: 2= two sided (A); U=true mean difference is greater (B); L=true mean difference is lower (C). Default is 2.

R code

```
status0<-blood_pressure$mmhg[blood_pressure$status==0]
status1<-blood_pressure$mmhg[blood_pressure$status==1]

t.test(status0,status1,mu=0,alternative="two.sided",
       var.equal=TRUE)
```

R output

```
t = -10.4679, df = 53, p-value = 1.660e-14
```

Remarks:

- mu=*value* is optional and defines the value μ_0 to test against. Default is 0.

- alternative=*"value"* is optional and defines the type of alternative hypothesis: "two.sided"= two sided (A); "greater"=true mean difference is greater (B); "less"=true mean difference is lower (C). Default is "two.sided".

2.2.3   Welch test

Description: Tests if two population means μ_1 and μ_2 differ less than, more than or by a value d_0.

Assumptions:
- Data are measured on an interval or ratio scale.

- Data are randomly sampled from two independent Gaussian distributions.

- Standard deviations σ_1 and σ_2 of the underlying Gaussian distributions are unknown and not necessarily equal; estimated through the population standard deviation of each sample.

Hypotheses:
(A) $H_0 : \mu_1 - \mu_2 = d_0$ vs $H_1 : \mu_1 - \mu_2 \neq d_0$
(B) $H_0 : \mu_1 - \mu_2 \leq d_0$ vs $H_1 : \mu_1 - \mu_2 > d_0$
(C) $H_0 : \mu_1 - \mu_2 \geq d_0$ vs $H_1 : \mu_1 - \mu_2 < d_0$

Test statistic:
$$T = [(\overline{X}_1 - \overline{X}_2) - d_0] \bigg/ \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\text{with } s_j = \sqrt{\frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_i - \overline{X}_j)^2}, \, j = 1, 2$$

Test decision:
Reject H_0 if for the observed value t of T
(A) $t < t_{\alpha/2,\nu}$ or $t > t_{1-\alpha/2,\nu}$
(B) $t > t_{1-\alpha,\nu}$
(C) $t < t_{\alpha,\nu}$

p-value:
(A) $p = 2 \, P(T \leq (-|t|))$
(B) $p = 1 - P(T \leq t))$
(C) $p = P(T \leq t))$

Annotations:
- This test is also known as a two-sample t-test or Welch–Satterthwaite test

- The test statistic T approximately follows a t-distribution with $\nu = \left(\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}\right)^2 \bigg/ \left(\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right)$ degrees of freedom [Bernard Welch (1947) and Franklin Satterthwaite (1946) approximation].

- $t_{\alpha,\nu}$ is the α-quantile of the t-distribution with ν degrees of freedom.

- William Cochran and Gertrude Cox (1950) proposed an alternative way to calculate critical values for the test statistic.

- The assumption of two underlying Gaussian distributions can be relaxed if the sample sizes of both samples are large. Usually sample sizes $n_1, n_2 \geq 25$ or 30 for both distributions are considered to be large enough.

Example: To test the hypothesis that the mean systolic blood pressures of healthy subjects (status=0) and subjects with hypertension (status=1) are equal, hence $d_0 = 0$. The dataset contains $n_1 = 25$ subjects with status 0 and $n_2 = 30$ with status 1 (dataset in Table A.1).

SAS code

```
proc ttest data=blood_pressure h0=0 sides=2 cochran;
 class status;
 var mmhg;
run;
```

SAS output

```
Method          Variances  DF       t Value  Pr > |t|
Satterthwaite   Unequal    50.886   -10.45   <.0001
Cochran         Unequal    .        -10.45   0.0001
```

Remarks:

- The optional command `cochran` forces SAS to calculate the p-value according to the Cochran and Cox approximation.

- h0=*value* is optional and defines the value μ_0 to test against. Default is 0.

- `sides`=*value* is optional and defines the type of alternative hypothesis: 2=two sided (A); U=true mean difference is greater (B); L=true mean difference is lower (C). Default is 2.

R code

```
status0<-blood_pressure$mmhg[blood_pressure$status==0]
status1<-blood_pressure$mmhg[blood_pressure$status==1]

t.test(status0,status1,mu=0,alternative="two.sided",
       var.equal=FALSE)
```

R output

```
t = -10.4506, df = 50.886, p-value = 2.887e-14
```

Remarks:

- mu=*value* is optional and defines the value μ_0 to test against. Default is 0.

- The option `var.equal=FALSE` enables the Welch test.

- `alternative="`*value*`"` is optional and indicates the type of alternative hypothesis: "two.sided"=two sided (A); "greater"=true mean difference is greater (B); "less"=true mean difference is lower (C). Default is "two.sided".

- By default R has no option to calculate the Cochran and Cox approximation.

2.2.4   Paired z-test

Description: Tests if the difference of two population means $\mu_d = \mu_1 - \mu_2$ differs from a value d_0 in the case that observations are collected in pairs.

Assumptions:
- Data are measured on an interval or ratio scale and randomly sampled in pairs (X_1, X_2).

- X_1 follows a Gaussian distribution with mean μ_1 and variance σ_1^2. X_2 follows a Gaussian distribution with mean μ_2 and variance σ_2^2. The covariance of X_1 and X_2 is σ_{12}.

- The standard deviation $\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$ of the differences $X_1 - X_2$ is known.

Hypotheses:
(A) $H_0 : \mu_d = d_0$ vs $H_1 : \mu_d \neq d_0$
(B) $H_0 : \mu_d \leq d_0$ vs $H_1 : \mu_d > d_0$
(C) $H_0 : \mu_d \geq d_0$ vs $H_1 : \mu_d < d_0$

Test statistic: $Z = \frac{\overline{D} - d_0}{\sigma_d} \sqrt{n}$ with $\overline{D} = \frac{1}{n} \sum_{i=1}^{n}(X_{1i} - X_{2i})$

Test decision: Reject H_0 if for the observed value z of Z
(A) $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$
(B) $z > z_{1-\alpha}$
(C) $z < z_\alpha$

p-value:
(A) $p = 2\Phi(-|z|)$
(B) $p = 1 - \Phi(z)$
(C) $p = \Phi(z)$

Annotations:
- The test statistic Z follows a standard normal distribution.

- z_α is the α-quantile of the standard normal distribution.

- The assumption of a Gaussian distribution can be relaxed if the distribution of the differences is symmetric.

Example: To test if the mean intelligence quotient increases by 10 comparing before training (IQ1) and after training (IQ2) (dataset in Table A.2). It is known that the standard deviation of the difference is 1.40. Note: Because we are interested in a negative difference of means of IQ1 − IQ2, we must test against $d_0 = -10$.

SAS code

```
* Calculate the difference of each observation;
data iq_diff;
 set iq;
 diff=iq1-iq2;
run;

* Calculate the mean and sample size;
proc means data=iq_diff mean;
 var diff;
 output out=ztest mean=mean_diff n=n_total;
run;

* Calculate test statistic and two-sided p-value;
data ztest;
 set ztest;
 d0=-10;            * Set difference to test;
 sigma_diff= 1.40; * Set standard deviation;
 format p_value pvalue.;
 z= sqrt(n_total)*((mean_diff-d0)/sigma_diff);
 p_value=2*probnorm(-abs(z));
run;

* Output results;
proc print;
 var z p_value;
run;
```

SAS output

```
    z        p_value
 -1.27775    0.2013
```

Remarks:

- There is no SAS procedure to calculate this test directly.

- The one-sided p-value for hypothesis (B) can be calculated with `p_value_B=1-probnorm(z)` and the p-value for hypothesis (C) with `p_value_C=probnorm(z)`.

R code

```
# Set difference to test;
d0<--10
# Set standard deviation of the difference
sigma_diff<-1.40

# Calculate the mean of the difference
mean_diff<-mean(iq$IQ1-iq$IQ2)

# Calculate the sample size
n_total<-length(iq$IQ1)

# Calculate test statistic and two-sided p-value
z<-sqrt(n_total)*((mean_diff-d0)/sigma_diff)
p_value=2*pnorm(-abs(z))

# Output results
z
p_value
```

R output

```
> z
[1] -1.277753
> p_value
[1] 0.2013365
```

Remarks:

- There is no basic R function to calculate the two-sample z-test directly.

- The one-sided p-value for hypothesis (B) can be calculated with `p_value_B=1-pnorm(z)` and the p-value for hypothesis (C) with `p_value_C=pnorm(z)`.

2.2.5 Paired t-test

Description: Tests if the difference of two population means $\mu_d = \mu_1 - \mu_2$ differs from a value d_0 in the case that observations are collected in pairs.

Assumptions:
- Data are measured on an interval or ratio scale and randomly sampled in pairs (X_1, X_2).

- X_1 follows a Gaussian distribution with mean μ_1 and variance σ_1^2. X_2 follows a Gaussian distribution with mean μ_2 and variance σ_2^2. The covariance of X_1 and X_2 is σ_{12}.

- The standard deviations are unknown. The standard deviation σ_d of the differences is estimated through the population standard deviation s_d of the differences.

Hypotheses: (A) $H_0 : \mu_d = d_0$ vs $H_1 : \mu_d \neq d_0$
(B) $H_0 : \mu_d \leq d_0$ vs $H_1 : \mu_d > d_0$
(C) $H_0 : \mu_d \geq d_0$ vs $H_1 : \mu_d < d_0$

Test statistic: $T = \dfrac{\overline{D}-d_0}{s_d}\sqrt{n}$ with $s_d = \sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(D_i - \overline{D})^2}$,

$\overline{D} = \dfrac{1}{n}\sum\limits_{i=1}^{n} D_i$ and $D_i = X_{1i} - X_{2i}, i = 1, \ldots, n$

Test decision: Reject H_0 if for the observed value t of T
(A) $t < t_{\alpha/2,n-1}$ or $t > t_{1-\alpha/2,n-1}$
(B) $t > t_{1-\alpha,n-1}$
(C) $t < t_{\alpha,n-1}$

p-value: (A) $p = 2\,P(T \leq (-|t|))$
(B) $p = 1 - P(T \leq t))$
(C) $p = P(T \leq t))$

Annotations:
- The test statistic T is t-distributed with $n - 1$ degrees of freedom.

- $t_{\alpha,n-1}$ is the α-quantile of the t-distribution with $n - 1$ degrees of freedom.

- The assumption of a Gaussian distribution can be relaxed if the distribution of the differences is symmetric.

Example: To test if the mean intelligence quotient increases by 10 comparing before training (IQ1) and after training (IQ2) (dataset in Table A.2). Note: Because we are interested in a negative difference of means of IQ1 − IQ2, we must test against $d_0 = -10$.

SAS code

```
proc ttest data=iq h0=-10 sides=2;
 paired iq1*iq2;
run;
```

SAS output

```
DF    t Value    Pr > |t|
19    -1.29      0.2141
```

Remarks:

- The command `paired` forces SAS to calculate the paired t-test. Do not forget the asterisk between the variable names.

- `h0=`*value* is optional and indicates the value μ_0 to test against. Default is 0.

- `sides=`*value* is optional and indicates the type of alternative hypothesis: 2=two sided (A); U=true mean is greater (B); L=true mean is lower (C). Default is 2.

R code

```
t.test(iq$IQ1,iq$IQ2,mu=-10,alternative="two.sided",
       paired=TRUE)
```

R output

```
t = -1.2854, df = 19, p-value = 0.2141
```

Remarks:

- The command `paired=TRUE` forces R to calculate the paired t-test.

- `mu=`*value* is optional and indicates the value μ_0 to test against. Default is 0.

- `alternative=`*"value"* is optional and indicates the type of alternative hypothesis: "two.sides"=two sided (A); "greater"=true mean is greater (B); "less"=true mean is lower (C). Default is "two.sided".

References

Cochran W.G. and Cox G.M. 1950 *Experimental Designs*. John Wiley & Sons, Ltd.

Satterthwaite F.E. 1946 An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110–114.

Welch B.L. 1947 The generalization of Student's problem when several different population variances are involved. *Biometrika* **34**, 28–35.

3

Tests on the variance

This chapter contains statistical tests on the variance of normal populations. In the one-sample case it is of interest whether the variance of a single population differs from some pre-specified value, where the mean value of the underlying Gaussian distribution may be known or unknown. SAS and R do not provide the user with ready to use procedures or functions for the resulting χ^2-tests. For the two-sample cases it must be distinguished between independent and dependent samples. In the former case an F-test and in the latter case a t-test is appropriate. The SAS procedure `proc ttest` provides a way to calculate the test for the two-sided hypothesis. We additionally show how the test can be performed for the one-sided hypothesis. In R the function `var.test` calculates the test for all hypotheses. In SAS and R there is no convenient way to calculate the t-test for dependent samples and we provide code for it. For k-sample variance tests (Levene test, Bartlett test) please refer to Chapter 17 which covers ANOVA tests.

3.1 One-sample tests

This section deals with the question, if the variance differs from a predefined value.

3.1.1 χ^2-test on the variance (mean known)

Description: Tests if a population variance σ^2 differs from a specific value σ_0^2.

Assumptions:
- Data are measured on an interval or ratio scale.
- Data are randomly sampled from a Gaussian distribution.
- The mean μ of the underlying Gaussian distribution is known.

Hypotheses: (A) $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$
(B) $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$
(C) $H_0 : \sigma^2 \geq \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$

Test statistic: $X^2 = \left[\sum_{i=1}^{n}(X_i - \mu)^2\right] \Big/ \sigma_0^2$

Test decision: Reject H_0 if for the observed value X_0^2 of X^2

(A) $X_0^2 < \chi_{\alpha/2,n}^2$ or $X_0^2 > \chi_{1-\alpha/2,n}^2$

(B) $X_0^2 > \chi_{1-\alpha,n}^2$

(C) $X_0^2 < \chi_{\alpha,n}^2$

p-value: (A) $p = 2 \ \min(P(X^2 \leq X_0^2), 1 - P(X^2 \leq X_0^2))$

(B) $p = 1 - P(X^2 \leq X_0^2)$

(C) $p = P(X^2 \leq X_0^2)$

Annotations:
- The test statistic X^2 is χ^2-distributed with n degrees of freedom.
- $\chi_{\alpha,n}^2$ is the α-quantile of the χ^2-distribution with n degrees of freedom.
- The test is very sensitive to violations of the Gaussian assumption, especially if the sample size is small [see Sheskin (2007) for details].

Example: To test the hypothesis that the variance of the blood pressures of a certain populations equals 400 (i.e., the standard deviation is 20) with known mean of 130 mmHg. The dataset contains 55 patients (dataset in Table A.1).

SAS code

```
*Calculate squared sum;
data chi01;
 set blood_pressure;
 mean0=130;      * Set the known mean;
 square_diff=(mmhg-mean0)**2;
run;

proc summary;
 var square_diff;
 output out=chi02 sum=sum_square_diff;
run;

* Calculate test-statistic and p-values;
data chi03;
 set chi02;
 format p_value_A p_value_B p_value_C pvalue.;
 df=_FREQ_;
 sigma0=20;      * Set std under the null hypothesis;
 chisq=sum_square_diff/(sigma0**2);
 * p-value for hypothesis (A);
 p_value_A=2*min(probchi(chisq,df),1-probchi(chisq,df));
 * p-value for hypothesis (B);
 p_value_B=1-probchi(chisq,df);
```

```
 * p-value for hypothesis (C);
 p_value_C=probchi(chisq,df);
run;

* Output results;
proc print;
 var chisq df p_value_A p_value_B p_value_c;
run;
```

SAS output

```
chisq   df    p_value_A    p_value_B    p_value_C
49.595  55     0.6390       0.6805       0.3195
```

Remarks:

- There is no SAS procedure to calculate this χ^2-test directly.

R code

```
mean0<-130 # Set known mean
sigma0<-20 # Set std under the null hypothesis

# Calculate squared sum;
sum_squared_diff<-sum((blood_pressure$mmhg-mean0)^2)

# Calculate test-statistic and p-values;
df<-length(blood_pressure$mmhg)
chisq<-sum_squared_diff/(sigma0^2)
# p-value for hypothesis (A)
p_value_A=2*min(pchisq(chisq,df),1-pchisq(chisq,df))
# p-value for hypothesis (B)
p_value_B=1-pchisq(chisq,df)
# p-value for hypothesis (C)
p_value_C=pchisq(chisq,df)

# Output results
chisq
df
p_value_A
p_value_B
p_value_C
```

R output

```
> chisq
[1] 49.595
> df
[1] 55
> p_value_A
```

```
[1] 0.6389885
> p_value_B
[1] 0.6805057
> p_value_C
[1] 0.3194943
```

Remarks:

- There is no basic R function to calculate this χ^2-test directly.

3.1.2 χ^2-test on the variance (mean unknown)

Description: Tests if a population variance σ^2 differs from a specific value σ_0^2.

Assumptions:
- Data are measured on an interval or ratio scale.
- Data are randomly sampled from a Gaussian distribution.
- The mean μ of the underlying Gaussian distribution is unknown.

Hypotheses:
(A) $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$

(B) $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$

(C) $H_0 : \sigma^2 \geq \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$

Test statistic: $X^2 = \left[(n-1)S^2\right]/\sigma_0^2$ with $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Test decision: Reject H_0 if for the observed value X_0^2 of X^2

(A) $X_0^2 < \chi_{\alpha/2,n-1}^2$ or $X_0^2 > \chi_{1-\alpha/2,n-1}^2$

(B) $X_0^2 > \chi_{1-\alpha,n-1}^2$

(C) $X_0^2 < \chi_{\alpha,n-1}^2$

p-value:
(A) $p = 2 \ \min(P(X^2 \leq X_0^2), 1 - P(X^2 \leq X_0^2))$

(B) $p = 1 - P(X^2 \leq X_0^2)$

(C) $p = P(X^2 \leq X_0^2)$

Annotations:
- The test statistic χ^2 is χ^2-distributed with $n-1$ degrees of freedom.
- $\chi_{\alpha,n-1}^2$ is the α-quantile of the χ^2-distribution with $n-1$ degrees of freedom.
- The test is very sensitive to violations of the Gaussian assumption, especially if the sample size is small (Sheskin 2007).

Example: To test the hypothesis that the variance of the blood pressures of a certain population equals 400 (i.e., the standard deviation is 20) with unknown mean. The dataset contains 55 patients (dataset in Table A.1).

SAS code

```
* Calculate sample std and sample size;
proc means data=blood_pressure std;
 var mmhg;
 output out=chi01 std=std_sample n=n_total;
run;

* Calculate test-statistic and p-values;
data chi02;
 set chi01;
 format p_value_A p_value_B p_value_C pvalue.;
 df=n_total-1;
 sigma0=20;    * Set std under the null hypothesis;
 chisq=(df*(std_sample**2))/(sigma0**2);
 * p-value for hypothesis (A);
 p_value_A=2*min(probchi(chisq,df),1-probchi(chisq,df));
 * p-value for hypothesis (B);
 p_value_B=1-probchi(chisq,df);
 * p-value for hypothesis (C);
 p_value_C=probchi(chisq,df);
run;

* Output results;
proc print;
 var chisq df p_value_A p_value_B p_value_c;
run;
```

SAS output

```
chisq   df   p_value_A   p_value_B   p_value_C
49.595  54    0.71039     0.64480     0.35520
```

Remarks:

- There is no SAS procedure to calculate this χ^2-test directly.

R code

```
# Calculate sample std and sample size;
std_sample<-sd(blood_pressure$mmhg)
n<-length(blood_pressure$mmhg)

# Set std under the null hypothesis
sigma0<-20

# Calculate test-statistic and p-values;
df=n-1
chisq<-(df*std_sample^2)/(sigma0^2)
# p-value for hypothesis (A)
p_value_A=2*min(pchisq(chisq,df),1-pchisq(chisq,df))
```

```
# p-value for hypothesis (B)
p_value_B=1-pchisq(chisq,df)
# p-value for hypothesis (C)
p_value_C=pchisq(chisq,df)

# Output results
chisq
df
p_value_A
p_value_B
p_value_C
```

R output

```
> chisq
[1] 49.595
> df
[1] 54
> p_value_A
[1] 0.7103942
> p_value_B
[1] 0.6448029
> p_value_C
[1] 0.3551971
```

Remarks:

- There is no basic R function to calculate this χ^2-test directly.

3.2   Two-sample tests

This section covers two-sample tests, which enable us to test if the variances of two populations differ from each other.

3.2.1   Two-sample F-test on variances of two populations

Description: Tests if two population variances σ_1^2 and σ_2^2 differ from each other.

Assumptions:
- Data are measured on an interval or ratio scale.
- Data are randomly sampled from two independent Gaussian distributions with standard deviations σ_1 and σ_2.

Hypotheses: (A) $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$

(B) $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs $H_1 : \sigma_1^2 > \sigma_2^2$

(C) $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs $H_1 : \sigma_1^2 < \sigma_2^2$

Test statistic: $F = S_1^2 \Big/ S_2^2$ with $S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_{ji} - \overline{X}_j)^2, \quad j = 1, 2$

Test decision: Reject H_0 if for the observed value F_0 of F

(A) $F_0 < f_{\alpha/2;n_1-1,n_2-1}$ or $F_0 > f_{1-\alpha/2;n_1-1,n_2-1}$

(B) $F_0 > f_{1-\alpha;n_1-1,n_2-1}$

(C) $F_0 < f_{\alpha;n_1-1,n_2-1}$

p-value: (A) $p = 2 \, \min(P(F \leq F_0), 1 - P(F \leq F_0))$

(B) $p = 1 - P(F \leq F_0))$

(C) $p = P(F \leq F_0))$

Annotations:
- The test statistic F is F_{n_1-1,n_2-1}-distributed.
- $f_{\alpha;n_1-1,n_2-1}$ is the α-quantile of the F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.
- The test is very sensitive to violations of the Gaussian assumption.

Example: To test the hypothesis that the variances of the systolic blood pressure of healthy subjects (status=0) and subjects with hypertension (status=1) are equal. The dataset contains $n_1 = 25$ subjects with status 0 and $n_2 = 30$ with status 1 (dataset in Table A.1).

SAS code

```
*** Variant 1 ***;
* Only for hypothesis (A);
proc ttest data=blood_pressure h0=0 sides=2;
 class status;
 var mmhg;
run;

*** Variant 2 ***;
* For hypotheses (A),(B), and (C);
* Calculate the two standard deviations and;
* sample size;
proc means data=blood_pressure std;
 var mmhg;
 by status;
 output out=ftest01 std=stdvalue n=n_total;
run;

* Output the std in two different datasets;
data ftest02 ftest03;
 set ftest01;
 if status=0 then output ftest02;
 if status=1 then output ftest03;
run;
* Rename std and sample size of the subjects with;
* status=0;
data ftest02;
 set ftest02;
 rename stdvalue=std_status0
        n_total=n_status0;
run;
```

```
* Rename std and sample size of subjects with;
* status=1;
data ftest03;
 set ftest03;
 rename stdvalue=std_status1
        n_total=n_status1;
run;

* Calculate test statistic p-values;
data ftest04;
 merge ftest02 ftest03;
 format p_value_A p_value_B p_value_C pvalue.;

* Calculate numerator and denominator of the;
* F-statistic;
 std_num=max(std_status0,std_status1);
 std_den=min(std_status0,std_status1);

* Calculate the appropriate degrees of freedom;
 if std_num=std_status0 then
   do;
    df_num=n_status0-1;
    df_den=n_status1-1;
   end;
  else
   do;
    df_num=n_status1-1;
    df_den=n_status0-1;
   end;

* Calculate the test-statistic;
 f=std_num**2/std_den**2;

* p-value for hypothesis (A);
 p_value_A=2*min(probf(f,df_num,df_den),
                    1-probf(f,df_num,df_den));
* p-value for hypothesis (B);
 p_value_B=1-probf(f,df_num,df_den);
* p-value for hypothesis (C);
 p_value_C=probf(f,df_num,df_den);
run;

* Output results;
proc print;
 var f df_num df_den p_value_A p_value_B p_value_C;
run;
```

SAS output

```
Variant 1
          Equality of Variances
 Method       Num DF   Den DF    F Value    Pr > F
Folded F       24       29        1.04      0.9180
```

```
Variant 2

  f     df_num  df_den  p_value_A  p_value_B  p_value_C
1.03634  24      29       0.9180      0.4590      0.5410
```

Remarks:

- Variant 1 calculates only the p-value for hypothesis (A) as `proc ttest` only includes this as additional information using the test statistic $F = \max(s_1^2, s_2^2) / \min(s_1^2, s_2^2)$.
- Variant 2 calculates p-values for all three hypotheses.
- In some situations SAS calculates an erroneous p-value with the variant 1. This occurs if the degree of freedom of the numerator is greater than the degree of freedom of the denominator and the test statistic F is between 1 and the median of the F-distribution. Details are given by Gallagher (2006). If this is the case, use either variant 2, or use the F-value which `proc ttest` provides and the formula of variant 2 for the two-sided p-value.

R code

```
status0<-blood_pressure$mmhg[blood_pressure$status==0]
status1<-blood_pressure$mmhg[blood_pressure$status==1]

var.test(status0,status1,alternative="two.sided")
```

R output

```
F = 1.0363, num df = 24, denom df = 29, p-value = 0.918
```

Remarks:

- `alternative=`*"value"* is optional and indicates the type of alternative hypothesis: "two.sides" (A); "greater" (B); "less" (C). Default is "two.sided".

3.2.2   t-test on variances of two dependent populations

Description: Tests if two population variances σ_1^2 and σ_2^2 differ from each other.

Assumptions:
- Data are measured on an interval or ratio scale and are randomly sampled in pairs (X_1, X_2).
- X_1 follows a Gaussian distribution with mean μ_1 and variance σ_1^2. X_2 follows a Gaussian distribution with mean μ_2 and variance σ_2^2.

Hypotheses: (A) $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$

(B) $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs $H_1 : \sigma_1^2 > \sigma_2^2$

(C) $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs $H_1 : \sigma_1^2 < \sigma_2^2$

Test statistic: $T = \left[\sqrt{(n-2)}(S_1^2 - S_2^2) \right] / \left[\sqrt{4(1-r^2)S_1^2 S_2^2} \right]$

$$\text{with} \quad S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ji} - \overline{X}_j)^2 \; \text{ for } j = 1, 2$$

$$\text{and} \quad r = \frac{\displaystyle\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2)}{\sqrt{\displaystyle\sum_{i=1}^{n}(X_{1i} - \overline{X}_1) \sum_{i=1}^{n}(X_{2i} - \overline{X}_2)}}.$$

Test decision: Reject H_0 if for the observed value t of T

(A) $t < t_{\alpha/2, n-2}$ or $t > t_{1-\alpha/2, n-2}$

(B) $t > t_{1-\alpha, n-2}$

(C) $t < t_{\alpha, n-2}$

p-value: (A) $p = 2\, P(T \leq (-|t|))$

(B) $p = 1 - P(T \leq t))$

(C) $p = P(T \leq t))$

Annotations: • The test statistic T is t-distributed with $n-2$ degrees of freedom.

• $t_{\alpha, n-2}$ is the α-quantile of the t-distribution with $n-2$ degrees of freedom.

• This test is very sensitive to violations of the Gaussian assumption (Sheskin 2007, pp. 754–755).

• Here, r denotes the correlation coefficient between X_1 and X_2.

Example: To test the hypothesis that the variance of intelligence quotients before training (IQ1) and after training (IQ2) stays the same. The dataset contains 20 subjects (dataset in Table A.2).

SAS code

```
* Calculate sample standard deviations;
* and sample size;
proc means data=iq std;
 var iq1;
 output out=std1 std=std1 n=n_total;
run;
```

```
proc means data=iq std;
 var iq2;
 output out=std2 std=std2 n=n_total;
run;

data ttest01;
 merge std1 std2;
run;

* Calculate correlation coefficient;
proc corr data=iq OUTP=corr01;
 var iq1 iq2;
run;

data corr02;
 set corr01;
 if _TYPE_='CORR' and _NAME_='IQ1';
 rename IQ2 = r;
 drop _TYPE_;
run;

data ttest02;
 merge ttest01 corr02;
run;

* Calculate test statistic and two-sided p-value;
data ttest03;
 set ttest02;
 format p_value pvalue.;
 df=n_total-2;
 t=((df**0.5)*(std1**2-std2**2))/
                  (4*(1-r**2)*(std1**2)*(std2**2));
 p_value=2*probt(-abs(t),df);
run;

* Output results;
proc print;
 var t df p_value;
run;
```

SAS output

```
     t         df    p_value
0.007821987    18    0.9938
```

Remarks:

- There is no SAS procedure to calculate this test directly.

- The one-sided p-value for hypothesis (B) can be calculated with `p_value_B=1-probt(t,df)` and the p-value for hypothesis (C) with `p_value_C=probt(t,df)`.

R code

```
# Calculate sample standard deviations
# and sample size
std1=sd(iq$IQ1)
std2=sd(iq$IQ2)
n_total<-length(iq$IQ1)

# Calculate correlation coefficient
r<-cor(iq$IQ1,iq$IQ2)

# Calculate test statistic and two-sided p-value
df<-n_total-2;
t<-(sqrt(df)*(std1^2-std2^2))/(4*(1-r^2)*std1^2*std2^2)
p_value=2*pt(-abs(t),df)

# Output results
t
df
p_value
```

R output

```
> t
[1] 0.007821987
> df
[1] 18
> p_value
[1] 0.993845
```

Remarks:

- There is no basic R function to calculate this test directly.

- The one-sided p-value for hypothesis (B) can be calculated with `p_value_B=1-pt(t,df)` and the p-value for hypothesis (C) with `p_value_C=pt(t,df)`.

References

Gallagher J. 2006 The F test for comparing two normal variances: correct and incorrect calculation of the two-sided p-value. *Teaching Statistics* **28**, 58–60.

Sheskin D.J. 2007 *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall.