What to do?

1. More on hypothesis test
   - Test on least squares residuals
   - Test and confidence intervals
2. Comparing two normal samples
   - Comparing standard deviations
   - The F-distribution
   - Testing equality of unit variances
   - Comparing means when variances are known
   - Comparing means when variances are common but unknown
   - Comparing means when variances are unknown
3. The power of a test
   - Errors of type I and II
   - Defining the power
   - Power of global test

Test on least squares residuals.

Least squares adjustment:

- Our observations(typically measurements of lengths and angles) are stored in the vector b. The measurement b_i has variance $\sigma_0^2 u_i$ and the measurements are independent. σ_0^2 is the unit variance and most often set to 1.
- Our unknowns (also called the elements) are stored in the vector x. It will typically be coordinates of points.
- Observation equation(linearized): $b - b_0 = A(x - x_0) - r$ with d redundants, i.e. $d = n - p$ where n is the number of observations(length of b) and p is the number of unknown elements(length of x).

Test on least squares residuals.

Least squares adjustment:

- Weight matrix: C is diagonal with $c_{ii} = \frac{1}{u_i}$.
- Normal matrix: $N = A^\top CA$
- Hat matrix: $H = AN^{-1}A^\top$
- Estimated residual vector: $\hat{r} = (HC - I)b$
- Variance factor: $s_0^2 = \frac{1}{d}\hat{r}^\top C\hat{r}$
- Variance on $\hat{r}_i$ is given by

$$\sigma_0^2 V_{ii} = \sigma_0^2(c_{ii}^{-1} - h_{ii})$$

where c_{ii}, h_{ii} are the diagonal elements in C, H.

Test on least squares residuals.

Consider the i'th residual $\hat{r}_i$ with variance $\sigma_0^2 V_{ii}$. Test:

$$H_0 : E(\hat{r}_i) = 0$$

Test statistics:

$$Z = \frac{\hat{r}_i}{\sigma_0 \sqrt{V_{ii}}}.$$

Two-sided test, i.e. alternative hypothesis $H_A :: E(\hat{r}_i) \neq 0$,
then the accept region is determined by fractiles from the normal distribution

$$[z_{\alpha/2}, z_{1-\alpha/2}].$$

If the global test is rejected, we may substitute σ_0 by the posterior estimate s_0 yielding a $t(d)$-test instead, i.e

$$A_\alpha = [t(d)_{\alpha/2}, t(d)_{1-\alpha/2}].$$

Test and confidence intervals

$H_0 : \theta = \theta_0$ is not rejected on significance level α

$\Updownarrow$

θ_0 is included in the $(1-\alpha)$-confidence interval for θ.

...... or in other words:

The $(1-\alpha)$-confidence interval for θ consists of the θ_0, where
$H_0 : \theta = \theta_0$ is not rejected on significance level α.

Testing $H_0 : \sigma_1 = \sigma_2$

$X_1, \ldots, X_m$ sample from $\mathcal{N}(\mu_1, \sigma_1^2)$ and
$Y_1, \ldots, Y_n$ sample from $\mathcal{N}(\mu_2, \sigma_2^2)$.
Test: $H_0 : \sigma_1 = \sigma_2.$
Test statistics:

$$V = S_1^2/S_2^2,$$

where s_1 and s_2 are posterior standard deviations determined with $d_1 = m - 1$ and $d_2 = n - 1$ redundants.
Two-sided test, i.e. alternative hypothesis $H_A : \sigma_1 \neq \sigma_2$,
then the accept region is

$$A_\alpha = [F(d_1, d_2)_{\alpha/2}, F(d_1, d_2)_{1-\alpha/2}],$$

where $F(d_1, d_2)_\beta$ is the β-quantile in the socalled **F-distribution** $F(d_1, d_2)$ with defrees of freedom (d_1, d_2).

F-distribution

The distribution of

$$V = \frac{Y_1/d_1}{Y_2/d_2},$$

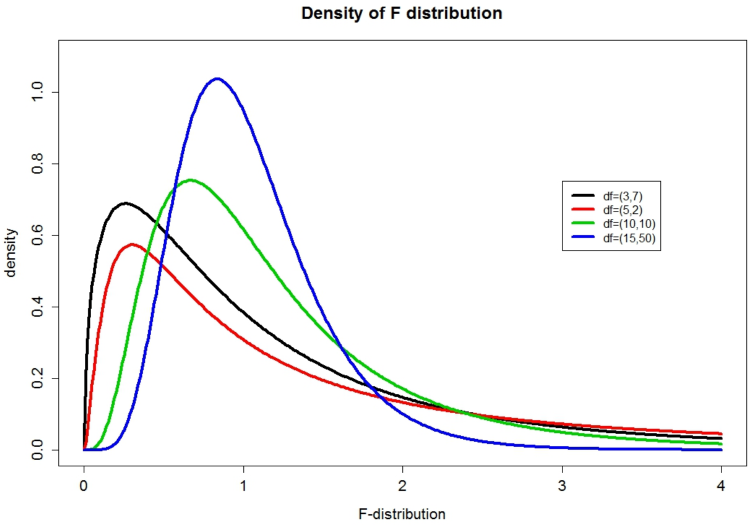
where Y_1 and Y_2 are independent and χ^2-distributed with d_1 og d_2 degrees of freedom, respectively, is called a *F-distribution* with (d_1, d_2) degrees of freedom. The density function is given as

$$f_V(v; d_1, d_2) = k_{d_1,d_2} \frac{v^{(d_2/2)-1}}{(d_1 v + d_2)^{(d_1+d_2)/2}}, \text{ for } v > 0.$$

It is relevant for comparison of two posteriori variances. Derived by R. A. Fisher around 1920. It holds that

$$E(V) = \frac{d_2}{d_2 - 2}.$$

Examples of F-distributions



Example on F-test

Observations $119, 112, 114$; earlier observations $110, 112, 109, 114$.
Same precision?
Test statistics:

$$f_{\text{obs}} = 13/4.9167 = 2.6441.$$

With significance level 5% the accept region is then

$$A_{0.05} = [F(2,3)_{0.025}, F(2,3)_{0.975}] = [0.0255, 16.04],$$

so it can not be documented that the standard deviation has changed.
Note that the accept region is very wide due to the low number of observations.

Python:

```
>>> from scipy.stats import f
>>> f.ppf([.025,.975],2,3)
array([  0.02553268,  16.04410643])
```

Testing equality of unit variances

We want to compare the results of two least squares adjustments, e.g. the positioning of an object at two different timepoints.
At timepoint i, $i = 1, 2$ let s_i^2 be the variancefactor , which estimates the unit variance σ_i^2 based on d_i redundants.
More specifically, we consider the null hypothesis

$$H_0 : \ \sigma_1 = \sigma_2$$

with the alternative

$$H_A : \ \sigma_1 \neq \sigma_2$$

Test statistics: $V = S_1^2 / S_2^2$
Accept region is

$$A_\alpha = [F(d_1, d_2)_{\alpha/2}, F(d_1, d_2)_{1-\alpha/2}],$$

where $F(d_1, d_2)_\beta$ is the β-quantile in $F(d_1, d_2)$.

Testing $H_0 : \mu_1 = \mu_2$ – known variances

$X_1, \ldots, X_m$ sample from $\mathcal{N}(\mu_1, \sigma_1^2)$ with σ_1 known and $Y_1, \ldots, Y_n$ sample from $\mathcal{N}(\mu_2, \sigma_2^2)$ with σ_2 known.
Test:

$$H_0 : \mu_1 = \mu_2.$$

Test statistics:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}},$$

Two-sided test, i.e. alternative hypothesis $H_A : \mu_1 \neq \mu_2$, then the accept region is

$$A_\alpha = [z_{\alpha/2}, z_{1-\alpha/2}].$$

where z_β is the β-quantile in $\mathcal{N}(0, 1)$.

Example: z-test.

Observations from earlier:
$119, 112, 114$ and $110, 112, 109, 114$,
standard deviations $\sigma_1 = 3$ and $\sigma_2 = 2$ known.
Test statistics:

$$z_{\text{obs}} = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} = \frac{115 - 111.25}{\sqrt{3^2/3 + 2^2/4}} = 1.875.$$

With 5% significance level it is contained in the accept region

$$A_\alpha = \texttt{norminv}([0.025\ 0.975]) = [-1.96, 1.96],$$

so we can not reject the null hypothesis, i.e. the mean has not changed.

Testing $H_0 : \mu_1 = \mu_2$ when $\sigma_1 = \sigma_2$ is unknown.

$X_1, \dots, X_m$ sample from $\mathcal{N}(\mu_1, \sigma_1^2)$ with σ_1 unknown and
$Y_1, \dots, Y_n$ sample from $\mathcal{N}(\mu_2, \sigma_2^2)$ with σ_2 unknown.

If we accept common standard deviation then combine s_1 and s_2

$$s^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2},$$

where $\frac{(m+n-2)S^2}{\sigma^2} \sim \chi^2(m+n-2)$.
Test: $H_0 : \mu_1 = \mu_2$.
Test statistic: $T = \frac{\bar{X}-\bar{Y}}{S\sqrt{1/m+1/n}}$,
Two-sided test, i.e. alternative hypothesis $H_A : \mu_1 \neq \mu_2$,
then the accept region is

$$A_\alpha = [t(m+n-2)_{\alpha/2}, t(m+n-2)_{1-\alpha/2}].$$

where $t(m+n-2)_\beta$ is β-quantile in $t(m+n-2)$.

Example: t-test.

Observations from earlier:
$119, 112, 114$ and $110, 112, 109, 114$,
standard deviation unknown, but accepted equal.
Common variance estimate:

$$s^2 = (2 \times 13 + 3 \times 4.9167)/5 = 8.15,$$

i.e. $s = \sqrt{8.15} = 2.855$ and the test statistics

$$t_{\mathrm{obs}} = \frac{\bar{x} - \bar{y}}{s\sqrt{1/m + 1/n}} = \frac{115 - 111.25}{2.855\sqrt{1/3 + 1/4}} = 1.72.$$

With 5% significance level it is contained in the accept region

$$A_\alpha = \mathtt{tinv}([0.025\ 0.975], 5) = [-2.57, 2.57],$$

so we do not reject the null hypothesis.

Testing $H_0 : \mu_1 = \mu_2$ with unknown variances.

$X_1, \ldots, X_m$ sample from $\mathcal{N}(\mu_1, \sigma_1^2)$ with σ_1 unknown and
$Y_1, \ldots, Y_n$ sample from $\mathcal{N}(\mu_2, \sigma_2^2)$ with
Test:

$$H_0 : \mu_1 = \mu_2.$$

It is natural to consider $\bar{X} - \bar{Y}$ and standardize this according to its variance $\sigma_1^2/m + \sigma_2^2/n$. We dont known the variance, but plug in the estimates to obtain the test statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/m + S_2^2/n}},$$

The distribution of T is complicated, but a good approximation - due to Welsch - is a t-distribution.

Unknown variances - continued.

The test statistic has an approximate t-distribution.
The degrees of freedom is determined as:

$$q = \frac{1}{a^2/(m-1) + (1-a)^2/(n-1)}$$

where

$$a = \frac{s_1^2/m}{s_1^2/m + s_2^2/n}$$

In case of a two-sided test, i.e. alternative hypothesis $H_A : \mu_1 \neq \mu_2$, we determine the acceptance region as

$$A_\alpha = [t(q)_{\alpha/2}, t(q)_{1-\alpha/2}].$$

where $t(q)_\beta$ is β-quantile in $t(q)$.

Example: t-test.

Observations from earlier:
$119, 112, 114$ and $110, 112, 109, 114$,
standard deviations unknown.
The test statistics

$$t_{\mathrm{obs}} = \frac{\bar{x} - \bar{y}}{\sqrt{s_1^2/m + s_2^2/n}} = \frac{115 - 111.25}{\sqrt{13/3 + 4.9167/4}} = 1.59.$$

and the degrees of freedom: $a = (13/3)(13/3 + 4.9167/4) = 0.779$ and $q = 1/(0.7792/2 + 0.2212/3) = 3.1279$.
With 5% significance the observed value 1.59 is contained in the accept region

$$A_\alpha = \mathtt{t.ppf}([0.025, 0.975], 3.1279) = [-3.11, 3.11],$$

so we do not reject the null hypothesis.

Error of type I and type II

In test of a hypothesis H_0 you can make two errors:

- Reject H_0, where H_0 is true - **Type I error**
- Accept H_0, where H_0 is false - **Type II error**

In test on significance level α, the probability for type I error is α, since the critical region K_α is determined by the requirement

$$\alpha = P(W \in K_\alpha \mid H_0) = P(\text{type I error}).$$

Type II error happens if H_0 is false, and $W \notin K_\alpha$, i.e.

$$P(W \notin K_\alpha \mid H_A) = P(\text{type II error}).$$

Power

A test has **great power,** if the probability of comitting a Type II error is small. The power is denoted by β and defined as

$$\beta = P(W \in K_\alpha \,|\, H_A),$$

i.e.

$$P(\text{type II error}) = 1 - \beta.$$

Typically β will **depend on the size** of the deviation from H_0, called δ. The function $\beta(\delta)$

$$\beta(\delta) = P(W \in K_\alpha \,|\, H_A(\delta)),$$

where $H_A(\delta)$ specifies H_A, is called the **power function**.

Example

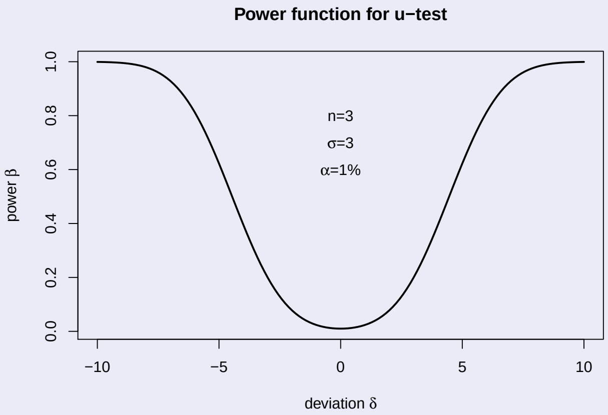
Observations: $119, 112, 114$ with standard deviation $\sigma = 3$.
Test $H_0 : \mu = 111$.
On significance level 1% we reject H_0, when

$$\bar{x} \notin [\mu_0 - z_{0.995} * \sigma/\sqrt{n}, \mu_0 + z_{0.995} * \sigma/\sqrt{n}] = [106.54, 115.46]$$

The power for $\mu = 111 + \delta$:

$$\begin{aligned}\beta(\delta) &= 1 - P(106.54 \leq \bar{X} \leq 115.46 \,|\, \mu = 111 + \delta) \\ &= 1 - \Phi\left(\frac{115.46 - 111 - \delta}{3/\sqrt{3}}\right) + \Phi\left(\frac{106.54 - 111 - \delta}{3/\sqrt{3}}\right) \\ &= 1 - \texttt{norm.cdf}\left((4.46 - \delta)/\sqrt{3}\right) + \texttt{norm.cdf}\left((-4.46 - \delta)/\sqrt{3}\right).\end{aligned}$$

A plot of the power function



Global test.

Least squares adjustment:

- Estimated residual vector: $\hat{r}$
- Weight matrix: C
- Number of redundants: d
- Prior unit variance: σ_0^2
- Design matrix: A
- Posterior unit variance: $s_0^2 = \frac{1}{d}\hat{r}^T C \hat{r}$

Test: $H_0 : E(s_0^2) = \sigma_0^2.$

Test statistics:

$$Y = \frac{dS_0^2}{\sigma_0^2}.$$

which in case of H_0 has a $\chi^2(d)$-distribution.

The power for the global test

The test statistic $Y = \frac{\hat{r}^\top C \hat{r}}{\sigma_0^2} \sim \chi^2(d)$

With significance level at e.g. 5% and $d=2$ redundants we reject H_0, when

$$y_{\text{obs}} \notin \texttt{chi2inv}([0.025\ 0.975], 2) = [0.051, 7.38].$$

To determine the power function, we shall specify the alternative, eg. an outlier of size δ.

More generally, let $e = E(r)$ be a vector of systematic errors, which under H_0 is the zero vector. Define

$$P = CA(A^T CA)^{-1} A^T C$$

One can show that the power function only depends on the size of $\lambda = (e^\top Ce - e^\top Pe)/\sigma_0^2$

Power for the global test — continued

In the example

- C is the identity matrix and $\sigma_0 = 3$.
- $A^\top = [1, 1, 1]$
- we assume a systematic error on the first measurement: $e = (\delta, 0, 0)$.

Hence

$$\lambda = (\delta^2 - \frac{1}{3} e^\top A A^\top e)/9 = \frac{2}{27}\delta^2$$

The power function can be calculated by means of the **non-central χ^2-distribution**, with density function:

$$f(x \mid d; \lambda) = \sum_{j=0}^{\infty} \frac{e^{-\frac{\lambda}{2}} \lambda^j}{2^j j!} \frac{e^{-\frac{x}{2}} x^{d/2-1}}{2^{d/2}\Gamma(\frac{d}{2} + j)}$$

where d is the number of redundants. This distribution is named `ncx2` in python.

A plot of the power function