Hypothesis testing

Maria Süveges

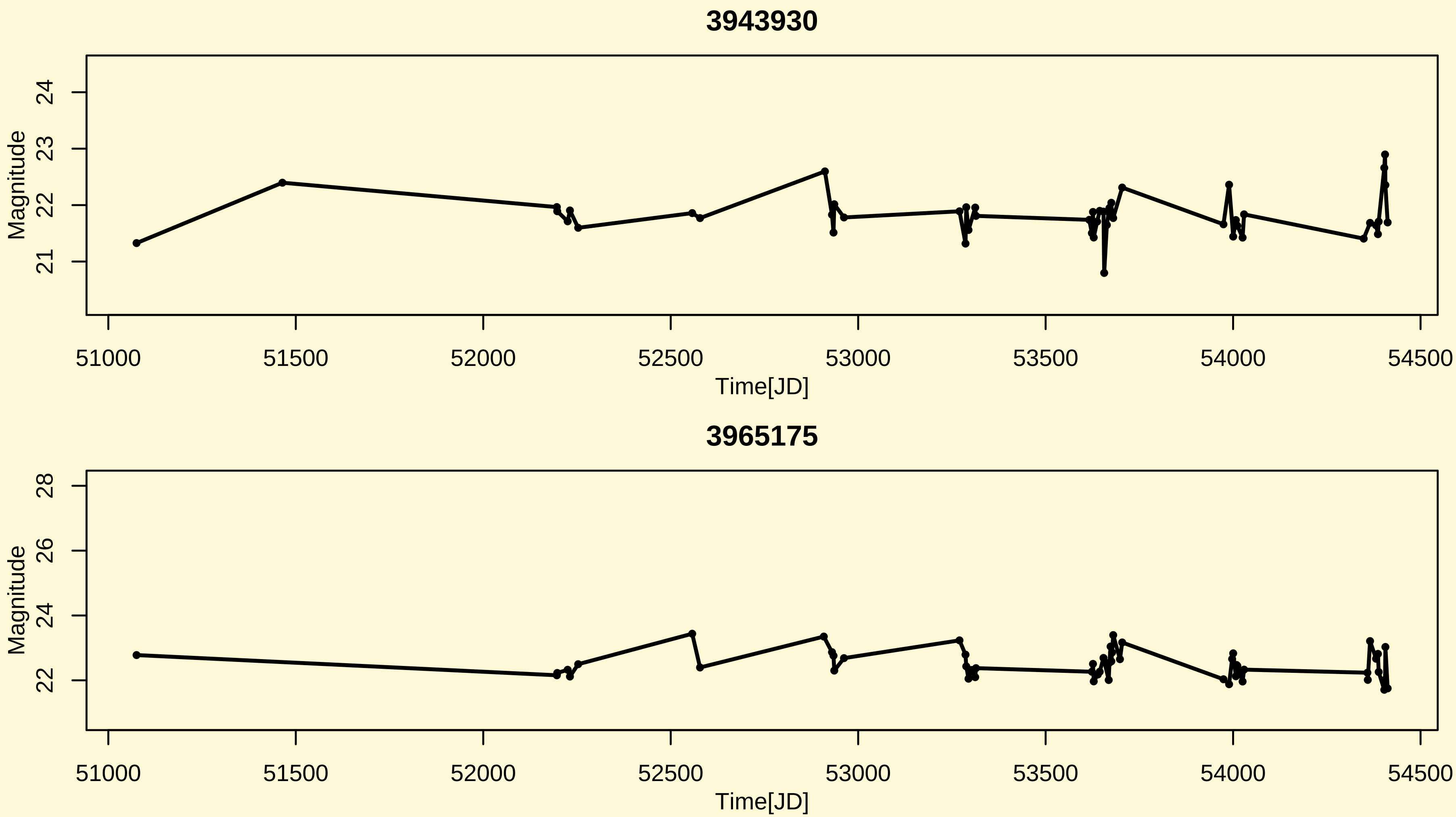
Laurent Eyer

September 22, 2010

Observatory of Geneva

Are these stars variable or not?

$Y_1, \ldots, Y_n$: time series of random variables (u-band magnitudes of any of two stars SDSS Stripe 82, $\lambda = 3551$Å)



Problems

If $Y_1, \ldots, Y_n$ are magnitudes of a star at $t_1, \ldots, t_n$, then

- eventual variability?
- mean magnitude?

Tough problem, because...

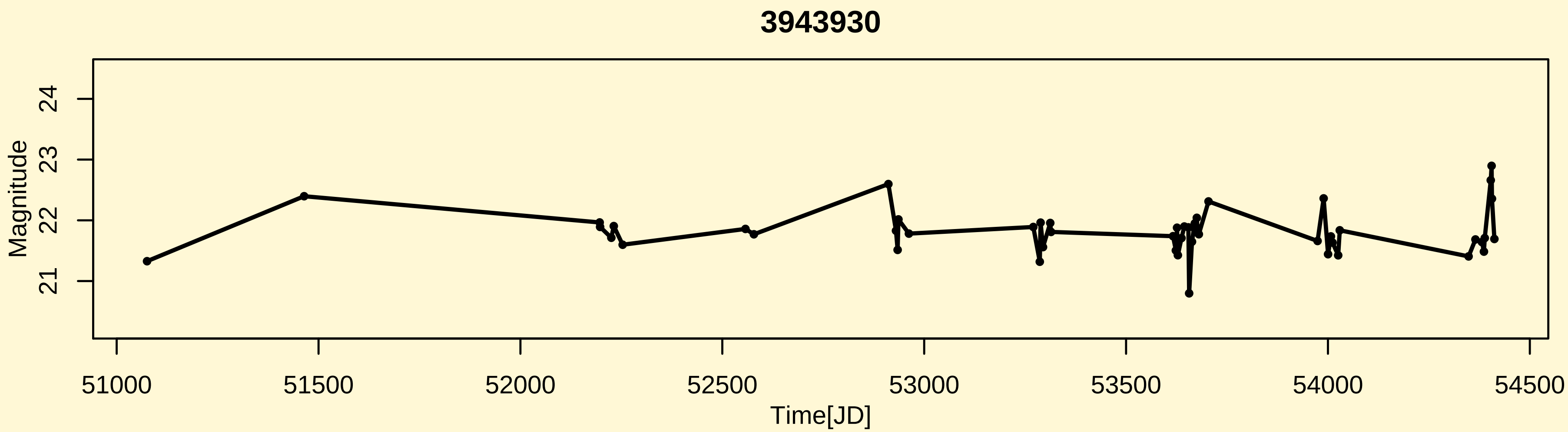
- ... random photon numbers on the detector;
- ... random electron numbers in the counter;
- ... errors of measurements: random and systematic instrumental errors, atmospheric effects, human mistakes;
- ... errors with many kinds of inter-dependence both on each other and on the true value to be measured;
- ... time series characteristics combined with irregular, but not completely random sampling;
- ... and so on.

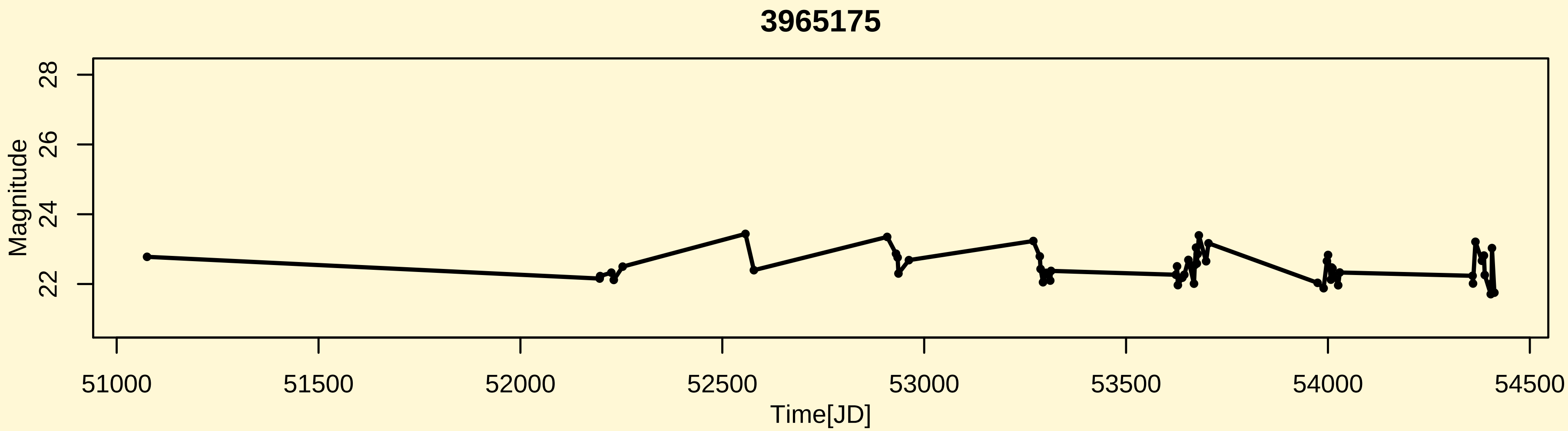
Statistical formulation

Put it in a very simple way:

1. Can the variations observed in the light curve be due entirely to the noise?

2. What is the mean magnitude of the stars?



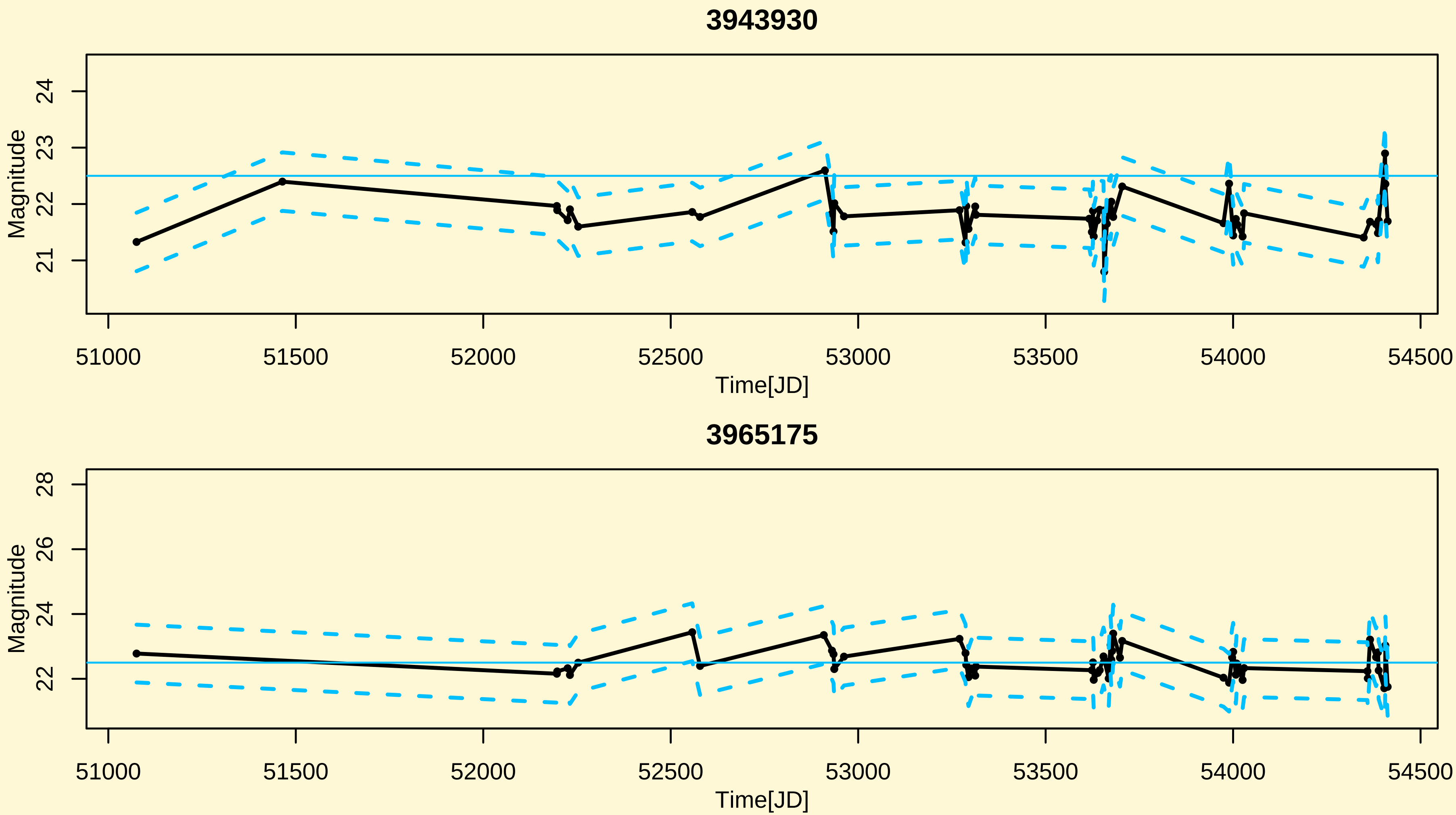


Statistical formulation

In a more statistical way:

1. Is the estimated standard error of the observations compatible with a given error σ_0?

2. Is the estimated mean compatible with some assumed (constant) value μ_0?



Sample quantities corresponding our questions:

1. The average

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

2. The empirical variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2.$$

We know from Laurent:

For $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$,

$$\begin{aligned}
\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} &\sim \mathcal{N}(0,1), \\
\frac{\bar{Y}-\mu}{S/\sqrt{n}} &\sim t_{n-1}, \\
\frac{(n-1)S^2}{\sigma^2} &\sim \chi^2_{n-1}.
\end{aligned}$$

Hypothesis test for the variance:

0. Make the fundamental assumptions. Here: $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

1. Formulate the null hypothesis and the alternative. Here:

$$H_0 \ : \ \sigma^2 = \sigma_0^2 \qquad \text{against} \qquad H_1 \ : \ \sigma^2 > \sigma_0^2.$$

2. Choose a test statistic that has a known distribution under H_0:

$$\xi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2}{\sigma_0^2},$$

so that $\xi^2 \sim \chi^2_{n-1}$, and calculate its value ξ^2_{obs} on the sample.

3. Fix a significance level α (often, $\alpha = 0.05$). Compute the p-value: $p = \Pr_{H_0}\{\xi^2 > \xi^2_{\text{obs}}\}$, or find the critical quantile $c_\alpha = \chi^2_{n-1}(1-\alpha)$.

4. Reject H_0 if $p < \alpha$ or equivalently, if $\xi^2_{\text{obs}} > \chi^2_{n-1}(1-\alpha)$.

Statistical formulation

Hypothesis test for the mean:

0. Make the fundamental assumptions. Here: $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

1. Formulate the null hypothesis and the alternative:

$$H_0 \;:\; \mu = \mu_0 \qquad \text{against} \qquad H_1 \;:\; \mu \neq \mu_0.$$

2. Choose a test statistic that has a known distribution under H_0:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \qquad \text{so that} \qquad Z \sim \mathcal{N}(0,1), \quad \text{if } \sigma \text{ can be taken as known, or}$$

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \qquad \text{so that} \qquad T \sim t_{n-1}, \quad \text{if not.}$$

Calculate the value T_{obs} or Z_{obs} on the sample.

3. Fix a significance level α.

Compute the p-value: $p = \mathrm{P}_{H_0}\{T > t_{\text{obs}}\}$, or find the critical quantiles $-c_{\alpha/2} = c_{1-\alpha/2} = t_{n-1}(1 - \alpha/2)$.

4. Reject H_0 if $p < \alpha/2$ or if $p > 1 - \alpha/2$; equivalently, if $t_{\text{obs}} > t_{n-1}(1 - \alpha/2)$ or $t_{\text{obs}} < t_{n-1}(\alpha/2)$.

Only looks simple...

The main problem:

To find a test statistic for which we fully know its distribution.

For our case, the test statistics are based on the iid normality of Y_i.

Most often exact null distributions cannot be found.

What can help: distributional convergence.

- Central Limit Theorem;
- other asymptotic convergence theorems (maximum likelihood estimators, periodogram value at a given frequency, deviance statistic for model comparison);
- convergence to distributions that cannot be analytically calculated in general (tests for equality of distributions).

Testing for the variance

$$H_0 \ : \ \sigma^2 = \sigma_0^2 \qquad \text{against} \qquad H_1 \ : \ \sigma^2 > \sigma_0^2.$$

$$\xi^2_{\text{obs}} = \frac{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}{\sigma_0^2} \quad \sim \quad \chi^2_{n-1}$$

Star 3965175: $\sigma_0^2 = 0.45^2$

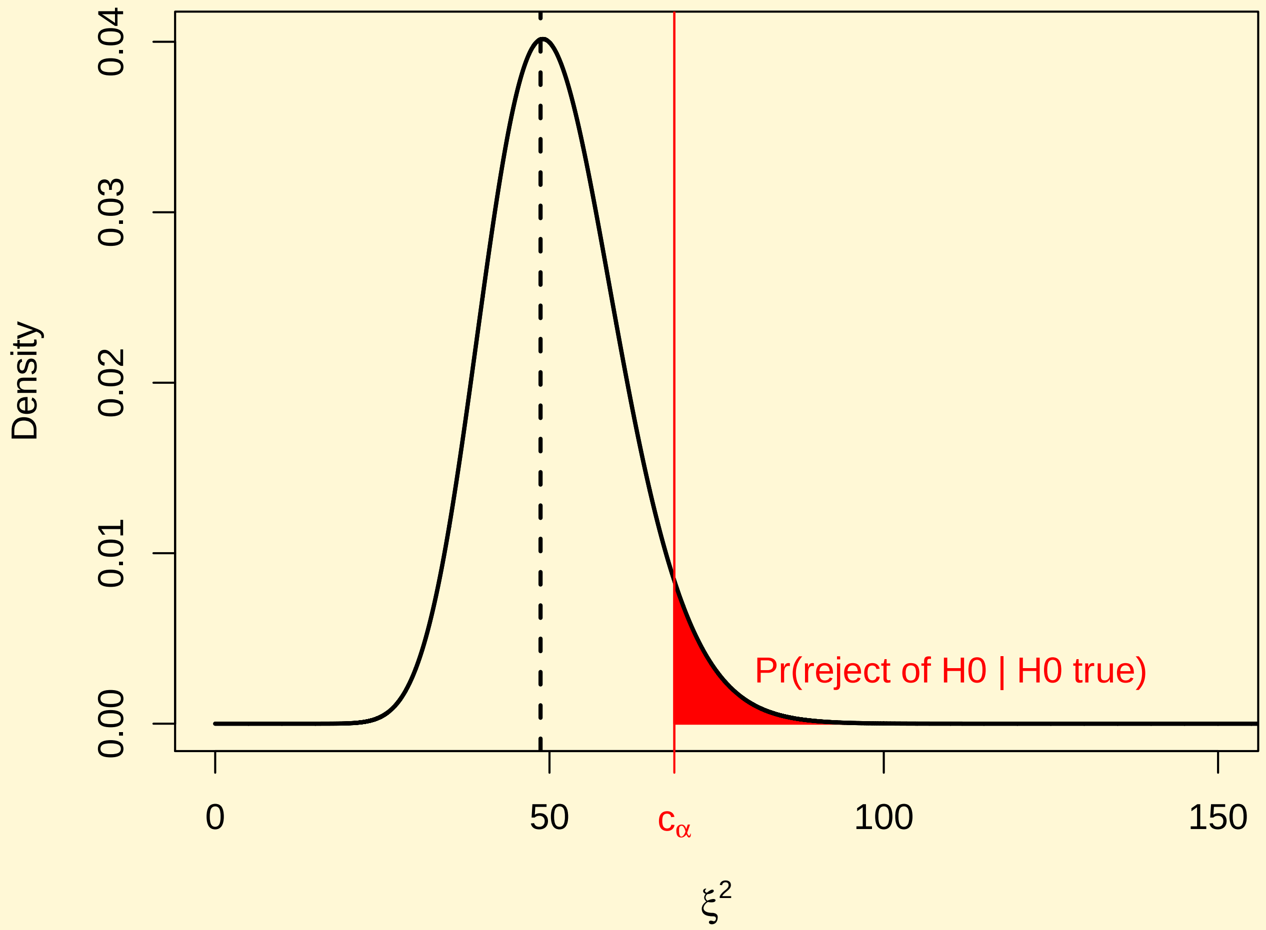
$n = 53$

$\chi^2_{n-1}(1-\alpha) = 69.83$

$\xi^2_{\text{obs}} = 48.93$

p-value $= 0.59$

H_0 not rejected.



$$H_0 \ : \ \sigma^2 = \sigma_0^2 \qquad \text{against} \qquad H_1 \ : \ \sigma^2 > \sigma_0^2.$$

$$\xi^2_{\text{obs}} = \frac{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2}{\sigma_0^2} \quad \sim \quad \chi^2_{n-1}$$

Star 3965175: $\sigma_0^2 = 0.45^2$

$n = 53$

$\chi^2_{n-1}(1-\alpha) = 69.83$

$\xi^2_{\text{obs}} = 48.93$

p-value $= 0.59$

H_0 not rejected.

Star 3943930: $\sigma_0^2 = 0.26^2$

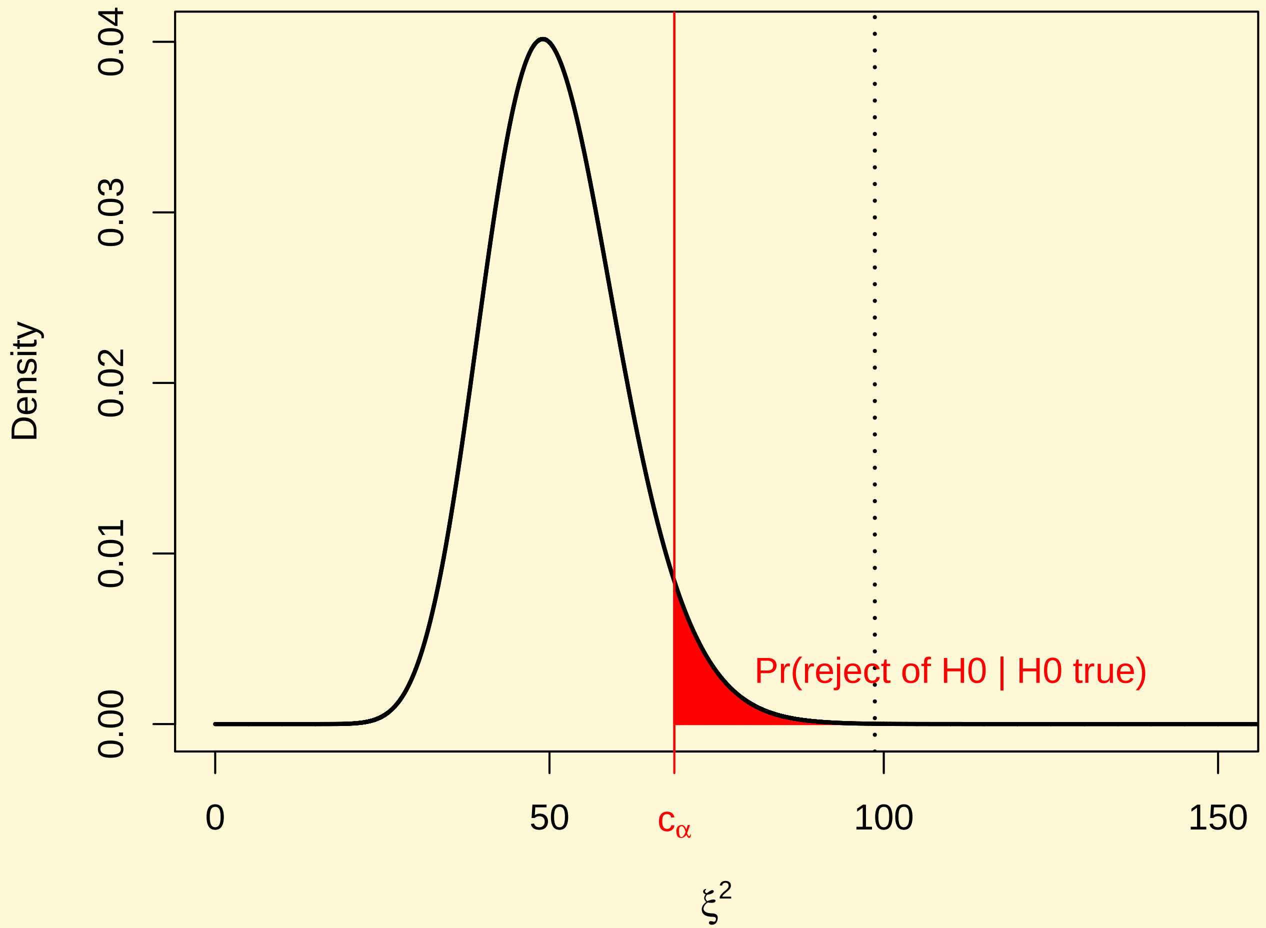
$n = 52$

$\chi^2_{n-1}(1-\alpha) = 68.67$

$\xi^2_{\text{obs}} = 98.66$

p-value $= 4.9 \times 10^{-5}$

H_0 rejected.



Testing for the mean I.

Star 3865175:

$$H_0 \;:\; \mu = 22.5 \qquad \text{against} \qquad H_1 \;:\; \mu \neq 22.5.$$

$$Z = \frac{\bar{X} - 22.5}{\sigma_0/\sqrt{n}} \,\Big|\, H_0 \quad \sim \quad \mathcal{N}(0,1)$$

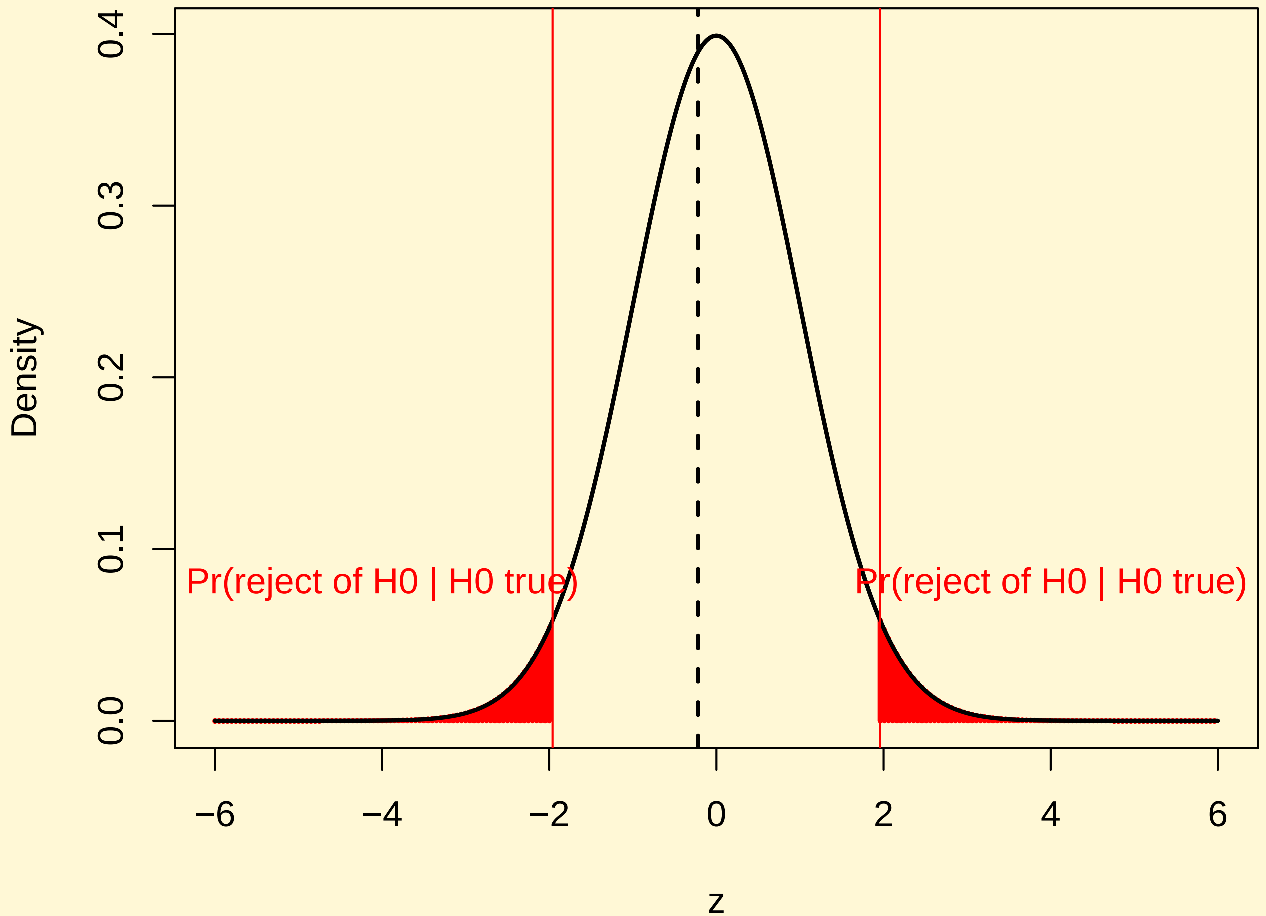
$\Phi^{-1}(1-\alpha/2) = 1.96$

$n = 53$

$Z_{\text{obs}} = -0.22$

p-value $= 0.59$

H_0 not rejected.

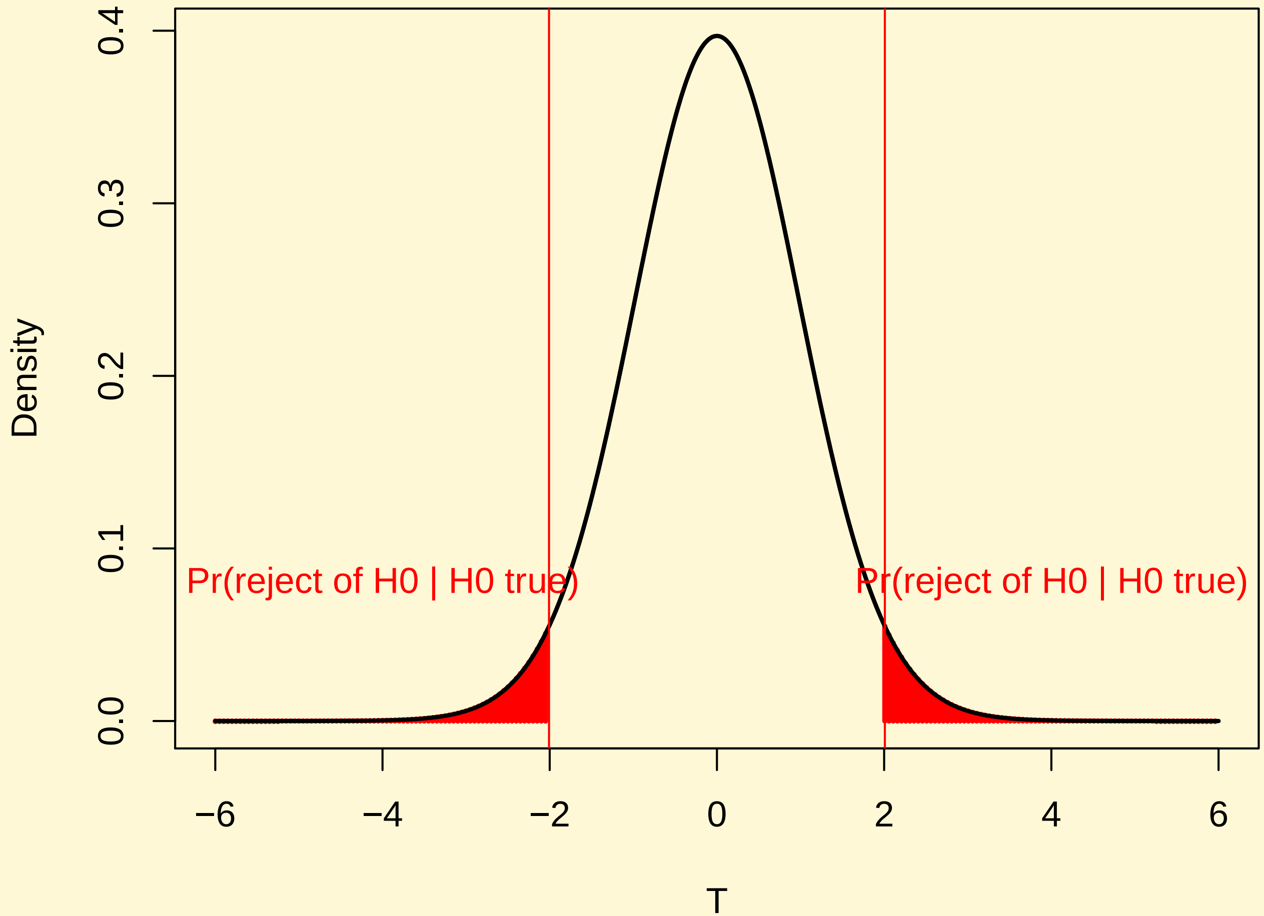


Testing for the mean II.

Star 3943930:

$$H_0 \ : \ \mu = 22.5 \qquad \text{against} \qquad H_1 \ : \ \mu \neq 22.5.$$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad \sim \quad t_{n-1}$$



$n = 52$

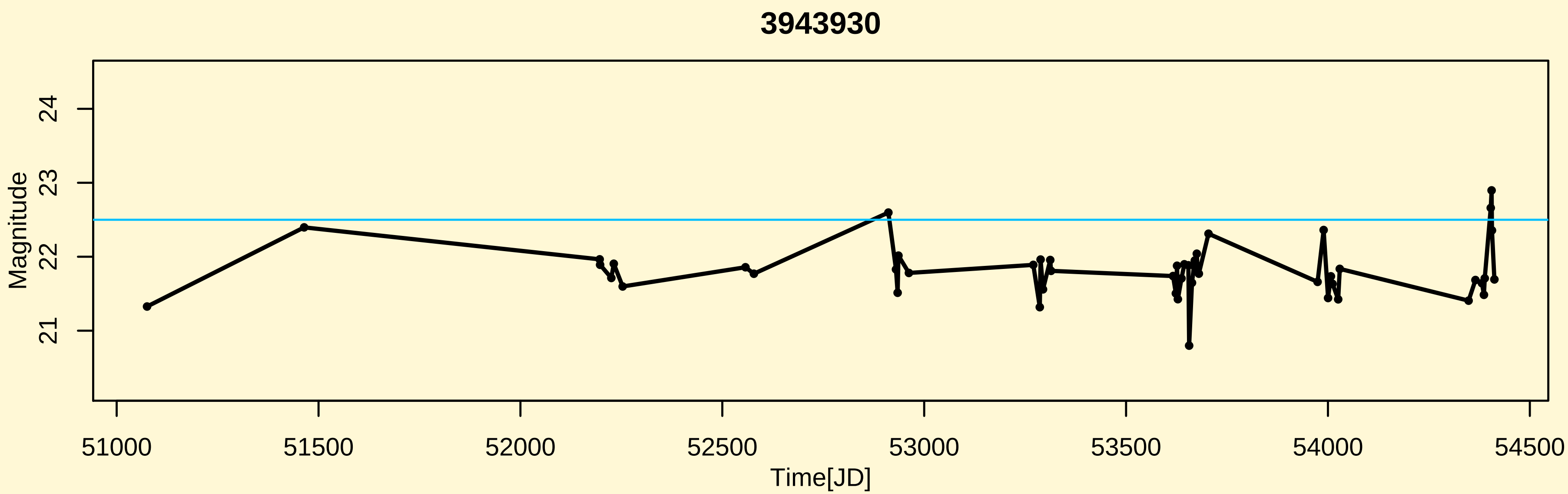
$t_{n-1}(1 - \alpha/2) = 2.01$

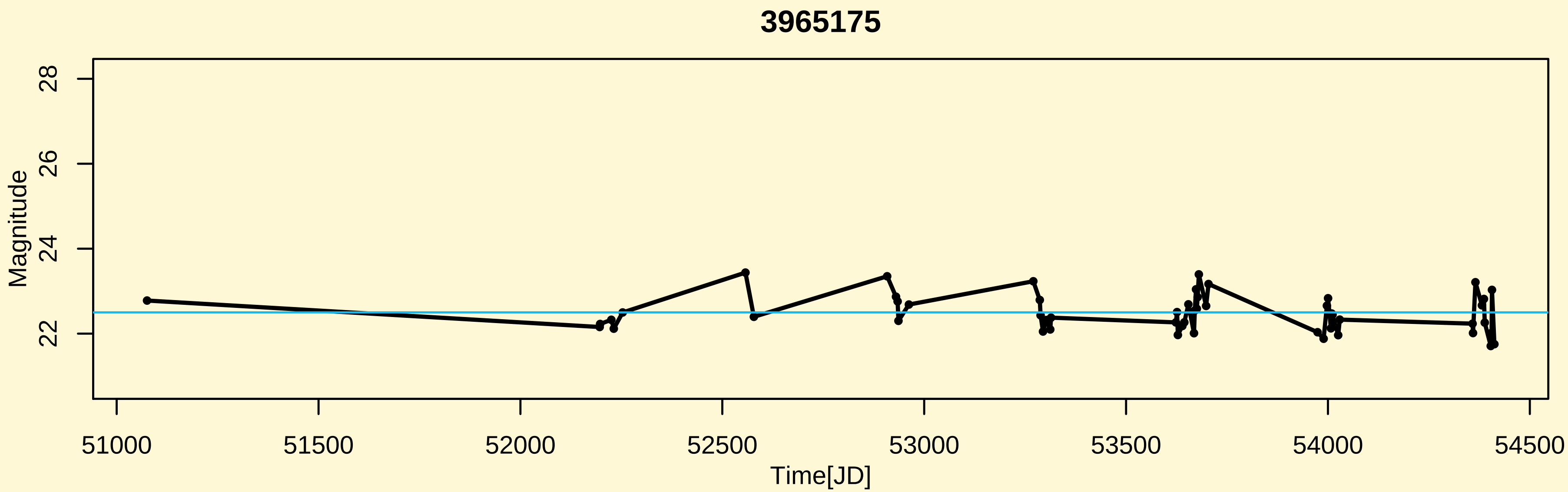
$t_{\text{obs}} = -13.69$

p-value $= 1$

H_0 rejected.

Have a look at the stars...





Types of errors

Decision:

$$\text{Reject } H_0 \quad \Longleftrightarrow \quad \xi^2 > c_\alpha.$$

	H_0 not rejected	H_0 rejected
H_0 true	Correct decision	Type I error, probability α
H_1 true	Type II error, probability β	Correct decision

Probability of Type I error:

$$\alpha = \Pr\left\{\xi^2 > c_\alpha \mid H_0\right\}$$

Probability of Type II error:

$$\beta = \Pr\left\{\xi^2 \leq c_\alpha \mid H_1\right\}$$

α : size of the test

$1 - \beta$: power of the test

For simplicity: let now $Y_1, \ldots, Y_n$ n measurements of the magnitudes of a star, suppose the errors on the measurements are all equal, and suppose $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

Hypotheses, now both simple:

$$H_0 \ : \ \sigma^2 = \sigma_0^2 = 0.26 \qquad \text{against} \qquad H_1 \ : \ \sigma^2 = \rho_0^2 = 0.36.$$

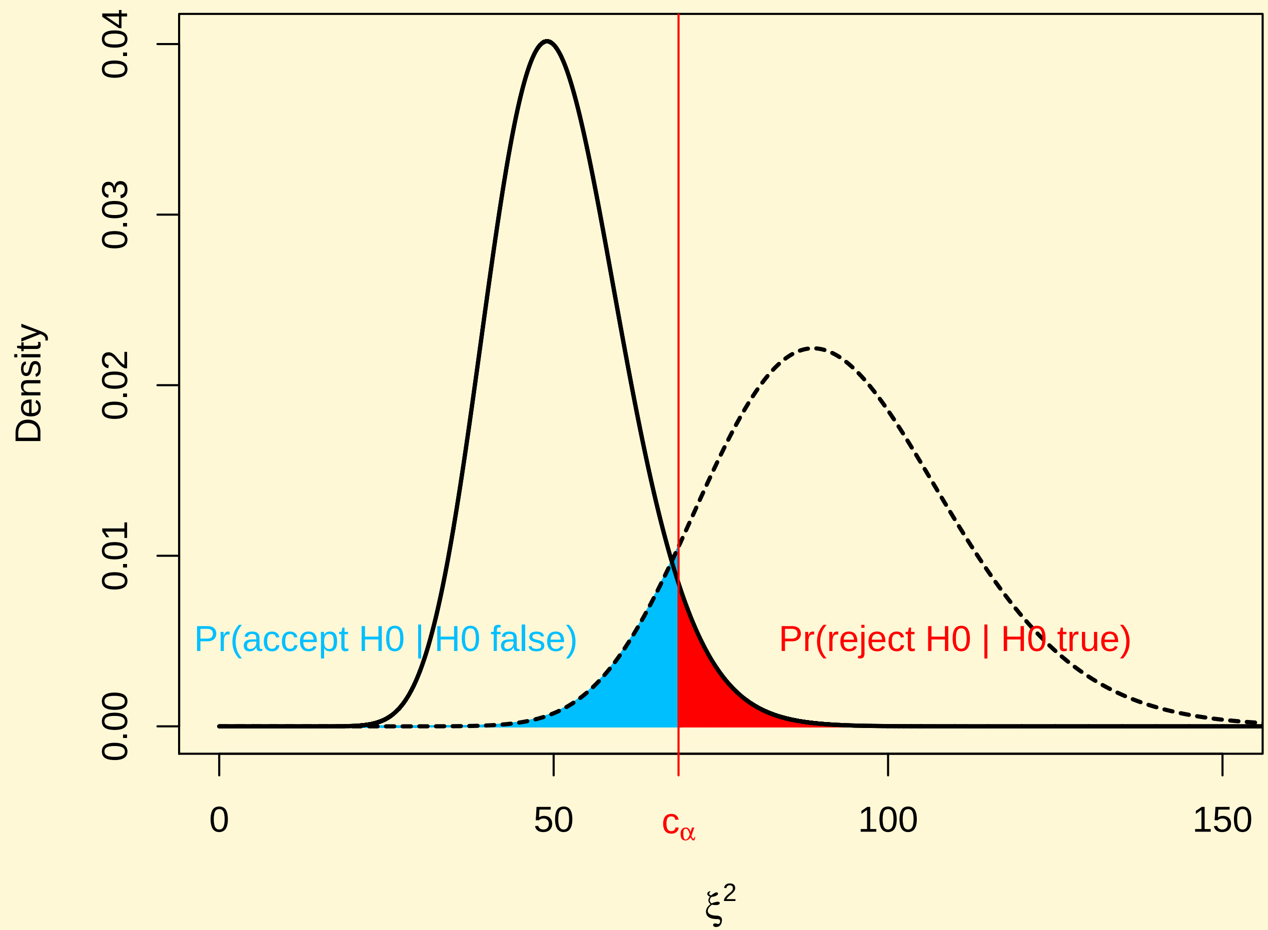
Test statistic:

$$\xi^2 = \frac{(n-1)S^2}{\sigma_0^2},$$

Under H_0,

$$\xi^2 \sim \chi^2_{n-1}.$$

Under H_1, a rescaled chi-squared.

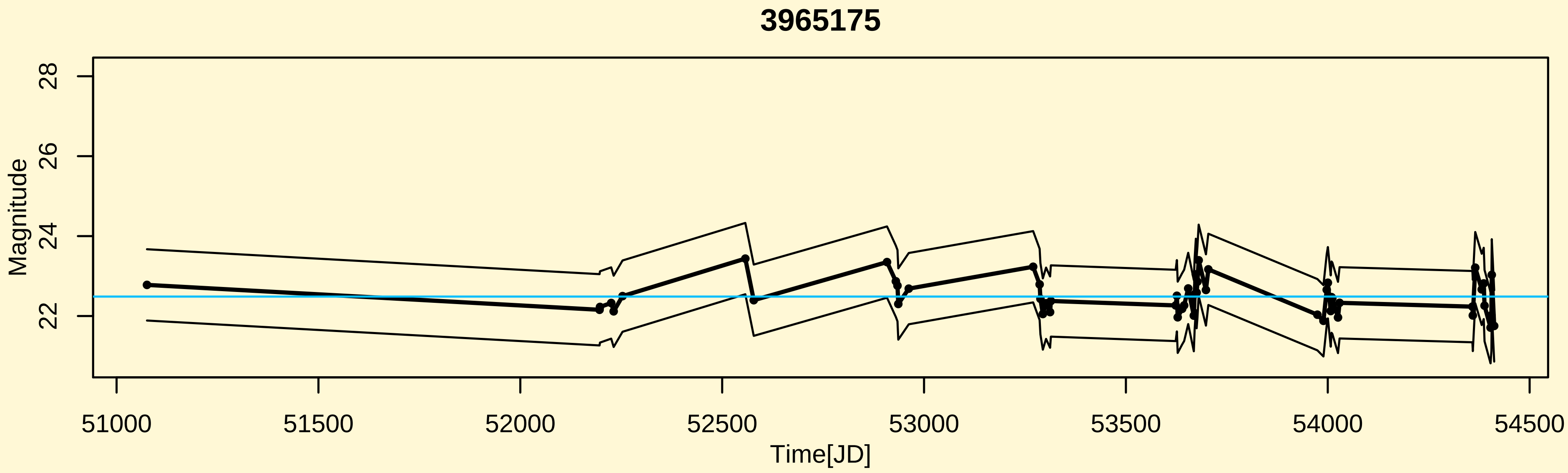


Types of errors

Attention: the type II error β cannot in general be exactly calculated, only if we have a simple alternative H_1 and we know the distribution of the test statistic under H_1.

- If $H_0 \ : \ \sigma^2 = \sigma_0^2$ against $H_1 \ : \ \sigma^2 = \rho_0^2$
  $\Longrightarrow \beta$ can be calculated;
- If $H_0 \ : \ \sigma^2 = \sigma_0^2$ against $H_1 \ : \ \sigma^2 > \sigma_0^2$
  $\Longrightarrow \beta$ cannot be calculated.

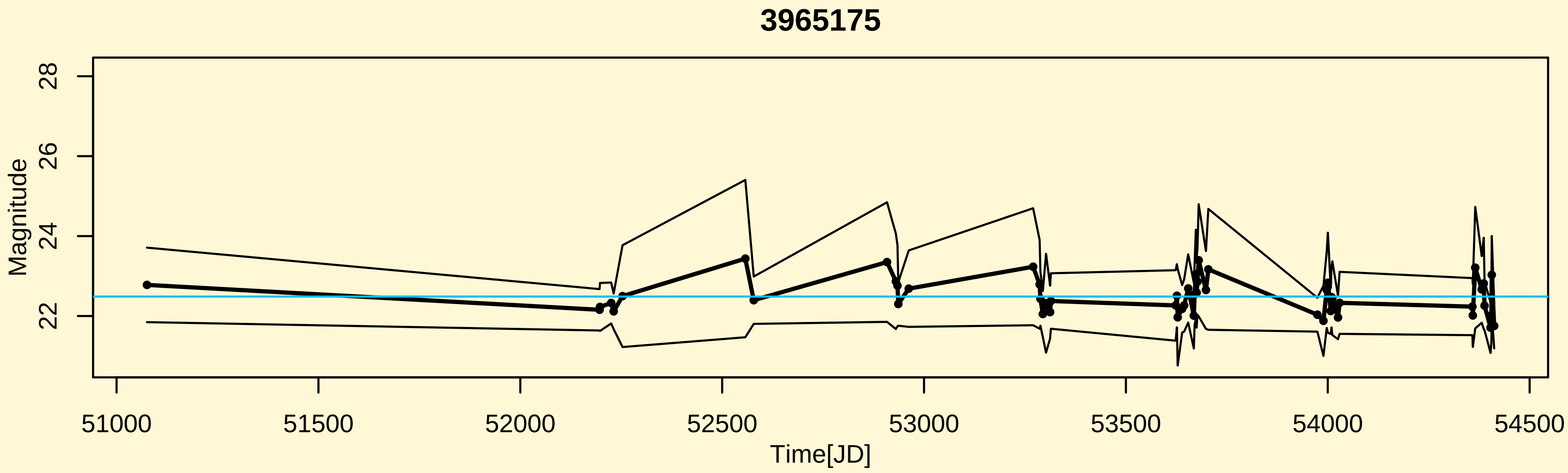
Complications



Fundamental assumption was:

$$Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

Complications



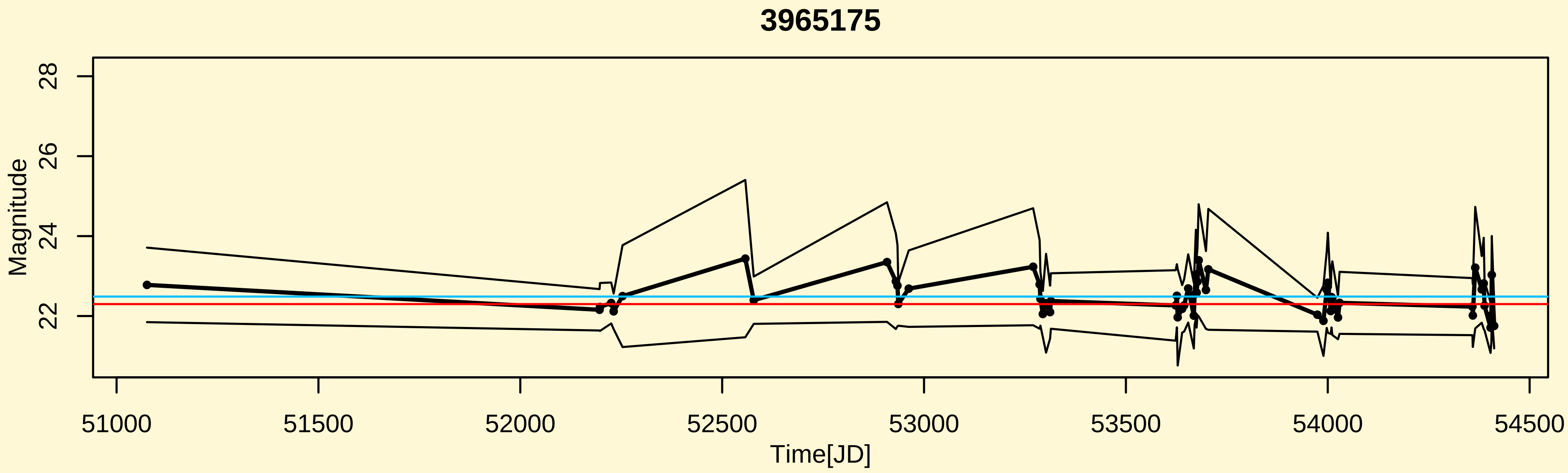
Fundamental assumption was:

$$Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

Not the case: Errors at different times are different!

$$Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu, \sigma_i^2).$$

Complications



Fundamental assumption was:

$$Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

Not the case: Errors at different times are different!

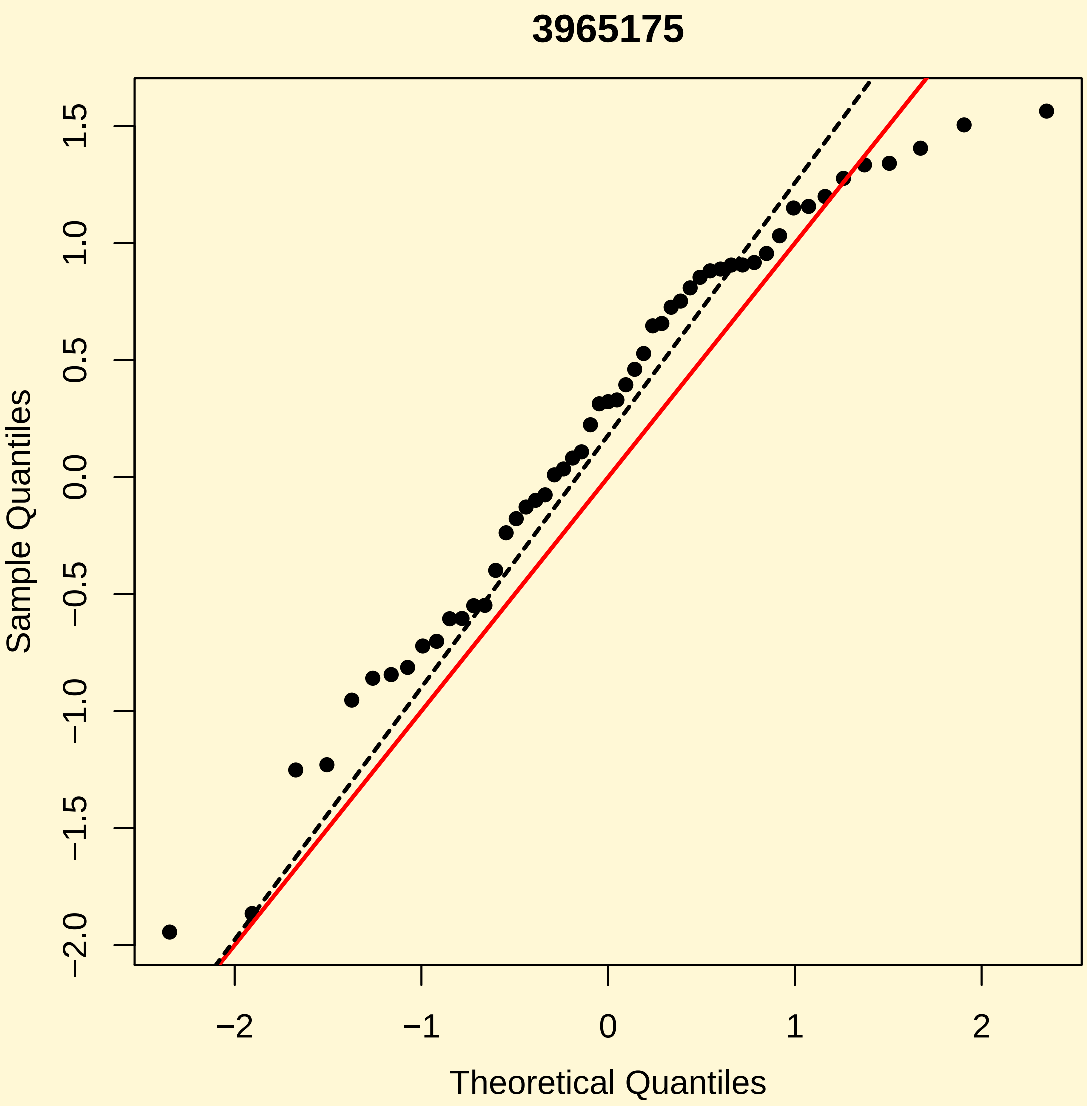
$$Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu, \sigma_i^2).$$

For the mean: maximum likelihood $\Longrightarrow$ weighted sample mean $\hat{\mu}$.

For the variance: standardize the observations by $Y_i^\star = (Y_i - \hat{\mu})/\sigma_i$, then as this is iid standard normal, compare $\sum_{i=1}^{n} Y_i^{\star 2}$ to a χ^2_{n-1}.

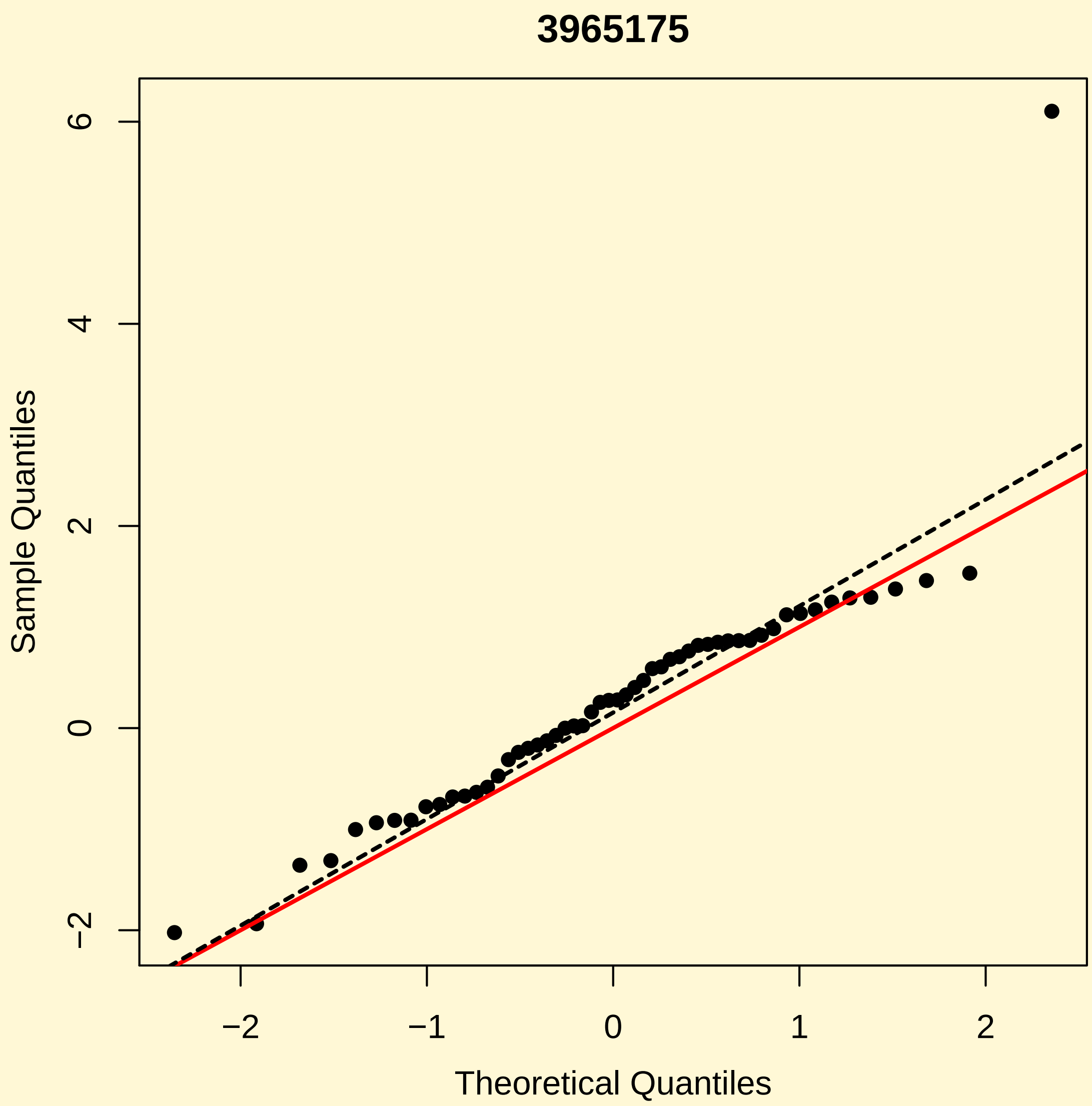
Complications

Are $Y_i^\star$ really standard normal?

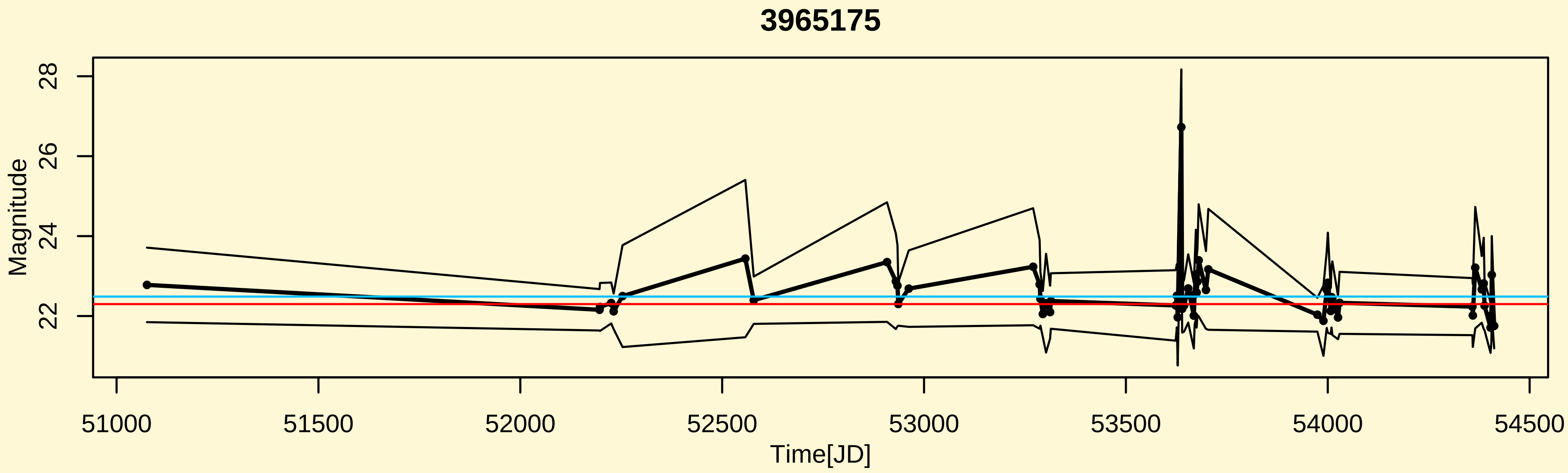


Complications

And this is not all....



Complications



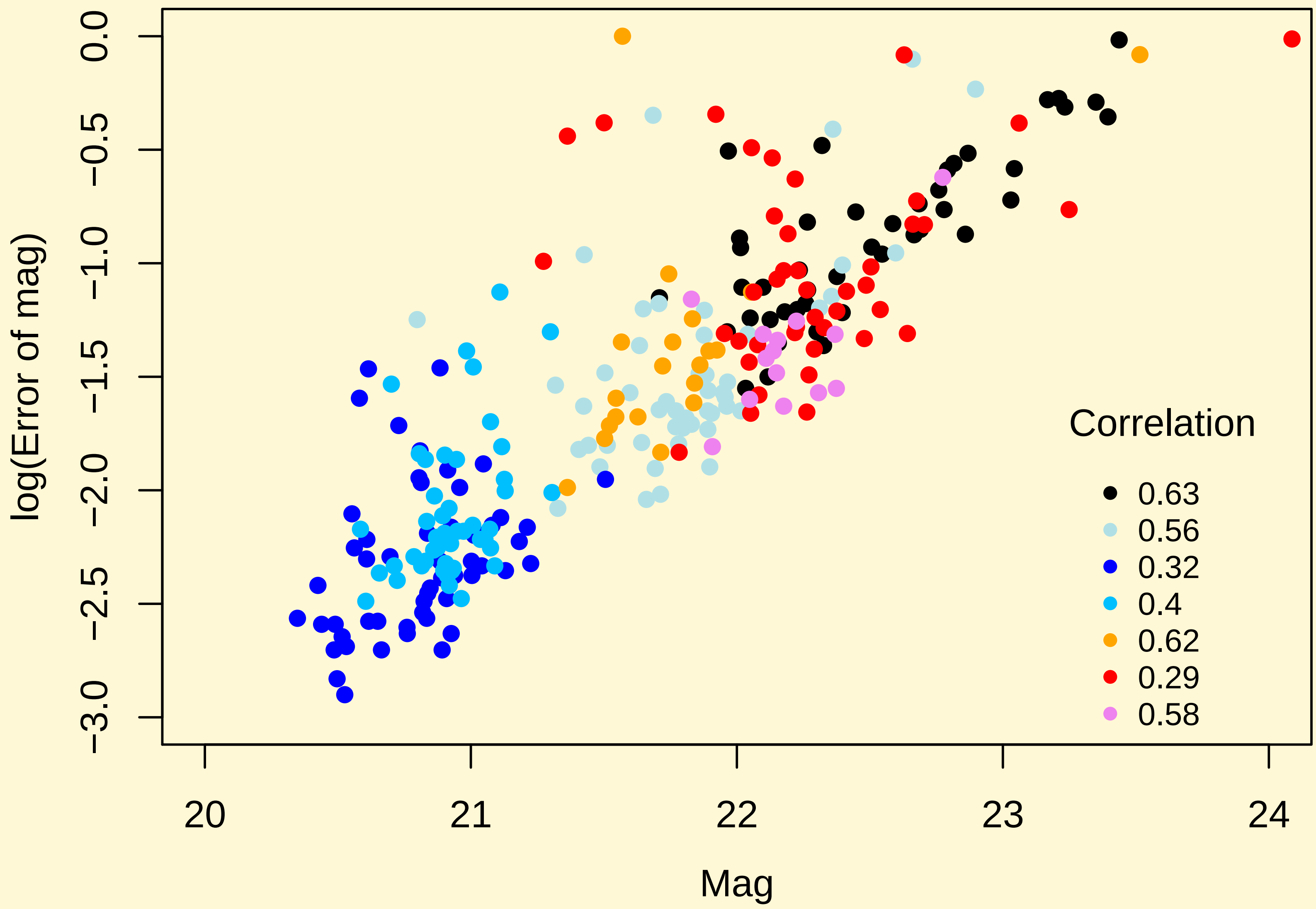
Outliers are still possible. Formally speaking:

$$Y_i \overset{\text{ind}}{\sim} \pi F_{1,i} + (1-\pi) F_{2,i}.$$

Solutions include:

- Remove outliers based on knowledge of experimental conditions;
- Apply robust statistical methods.

Complications



Remedy : maximum likelihood (should have some reasonable distributional assumption other than normality)

Crucial points

Precise formulation

What do I want to test?

$$H_0 \; : \; \ldots \qquad \text{against} \qquad H_1 \; : \; \ldots$$

Choice of test statistics

Corresponding to the precise formulation, and making the unavoidable assumptions and simplifications in order to have a fully known (asymptotic) distribution under H_0.

Clear on the underlying assumptions

What are the necessary conditions?

- Normality? Outliers?
  Check: QQ plots.
- Homogeneous errors?
  Check: plot the errors versus time (or your covariate).
- Independence of the errors and the observed quantity?
  Check: plot the errors versus the observed quantity.

What does the test tell us? Not that the star is variable: but only that the observations are more variable than our assumption about the noise.