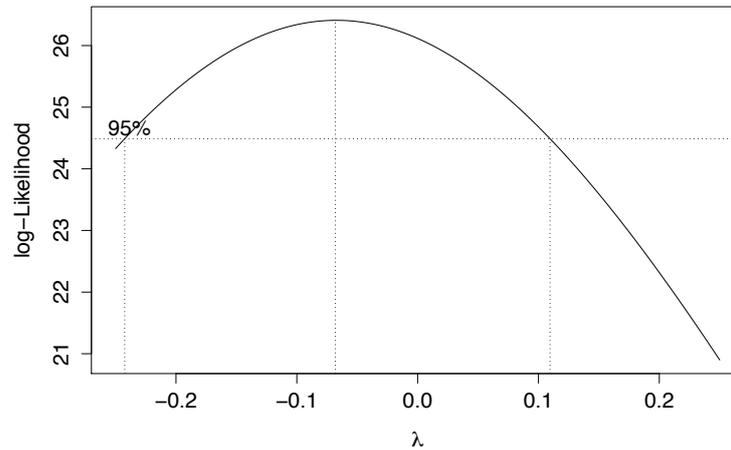


5K Statistical Modelling

The R command

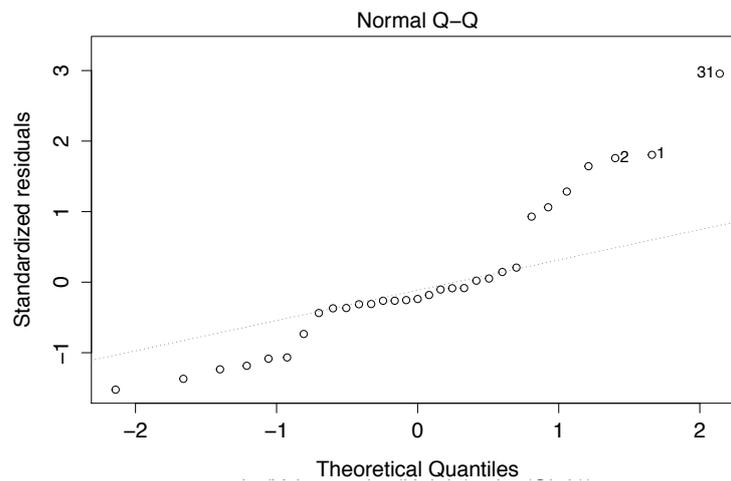
```
> boxcox(rainfall ~ month+elnino+month:elnino)
```

performs a Box–Cox transform of the response at several values of the parameter λ , and produces the following plot:



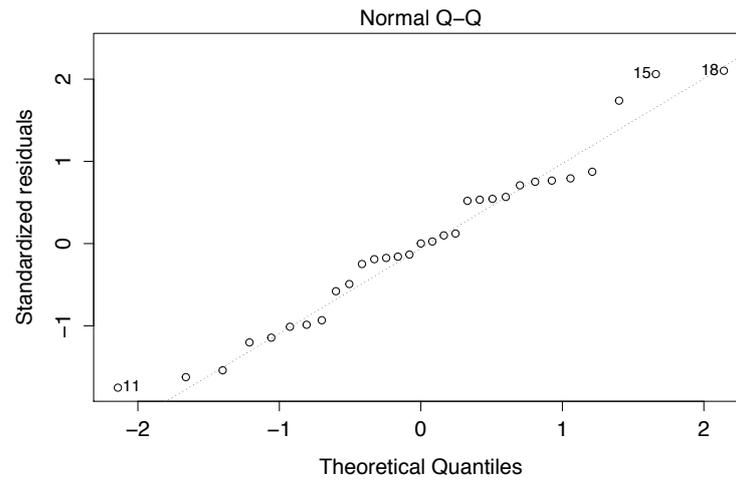
We fit two linear models and obtain the Q–Q plots for each fit, which are shown below in no particular order:

```
> fit.1 <- lm(rainfall ~ month+elnino+month:elnino)
> plot(fit.1,which=2)
> fit.2 <- lm(rainfall^-0.07 ~ month+elnino+month:elnino)
> plot(fit.2,which=2)
```



This question continues on the next page

5K Statistical Modelling (continued)



Define the variable on the y -axis in the output of `boxcox`, and match each Q-Q plot to one of the models.

After choosing the model `fit.2`, the researcher calculates Cook's distance for the i th sample, which has high leverage, and compares it to the upper 0.01-point of an $F_{p,n-p}$ distribution, because the design matrix is of size $n \times p$. Provide an interpretation of this comparison in terms of confidence sets for $\hat{\beta}$. Is this confidence statement exact?