# Assignment 3

## Ex. 3

***Consider the simple linear regression model:***
$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, N$
***where $\beta_0$ and $\beta_1$ are the unknown parameters. Assume that the $\epsilon_i$'s are iid $t$-distributed data with unknown degrees of freedom $\nu$.***

***(a) Write the loglikelihood function for the MLE estimation of the three unknowns parameters***.

Our model can be written as
$y \mid x \sim \tau(\beta_0 + \beta_1 x, \nu)$
where $\tau(\mu, \eta)$ is a (noncentral) Student's $t$-distribution centered in $\mu$. The loglikelihood $l$ of our model is

$$l\left((\beta_0, \beta_1, \nu \mid y, x\right) = \log \prod_{i=1}^{N} f(y_i - \beta_0 - \beta_1 x_i \mid \nu) = \sum_{i=1}^{N} \log f(y_i - \beta_0 - \beta_1 x_i \mid \nu)$$

where $f(\cdot \mid \nu)$ is the density function of a standard Student's $t$-distibution with $\nu$ degrees of freedom, given by

$$f(t \mid \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

so the log density takes the form

$$l\left((\beta_0, \beta_1, \nu \mid y, x\right) = N \log\left(\frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2}, \frac{\nu}{2}\right)}\right) - \frac{\nu+1}{2} \sum_{i=1}^{N} \log\left(1 + \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\nu}\right)$$

In [1]:
```
function loglikelihood(β0, β1, v, x, y)
    N = length(y)
    df_term = N * log(1 / sqrt(v) / beta(0.5, 0.5v))
    error_term = -0.5(v + 1) * sum(log.(1 + (y - β0 - β1*x).^2 / v))
    return df_term + error_term
end
```

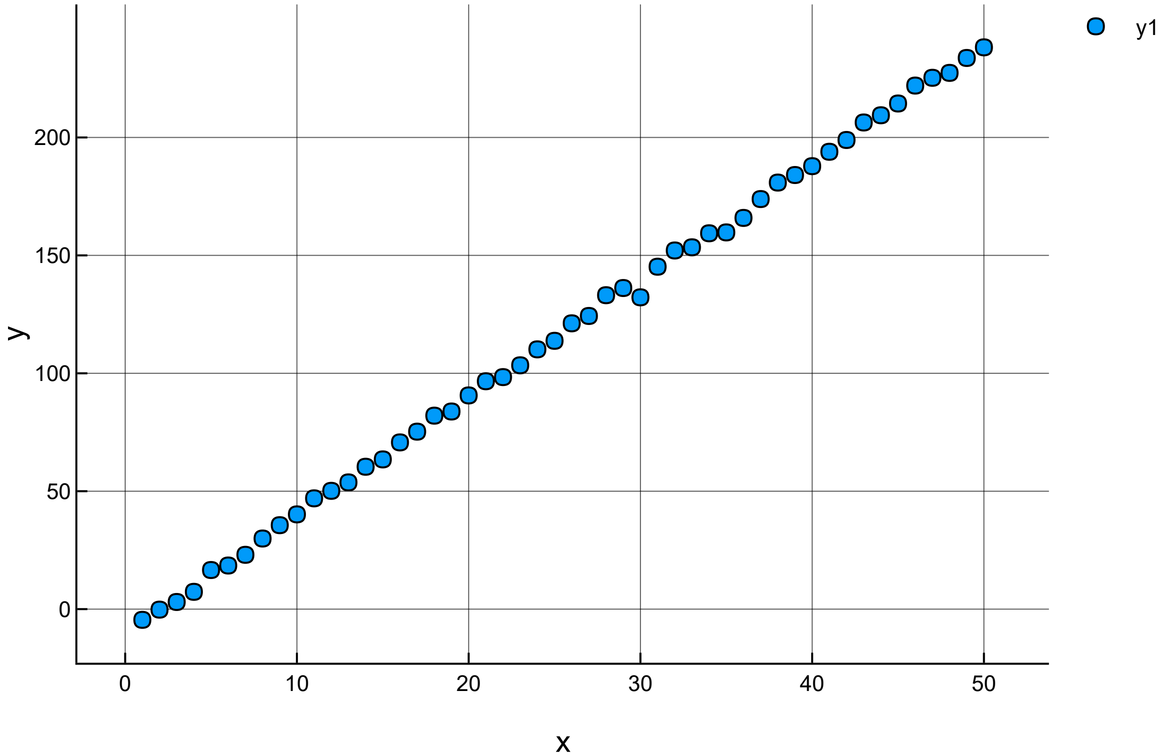
Out[1]:
```
loglikelihood (generic function with 1 method)
```

In [2]:
```
using DataFrames # read the data
using Plots # visualize the data with plots
using StatPlots # visualize the data with plots
```

In [3]:
```
tdata = readtable("../data/tdata.tsv");
```

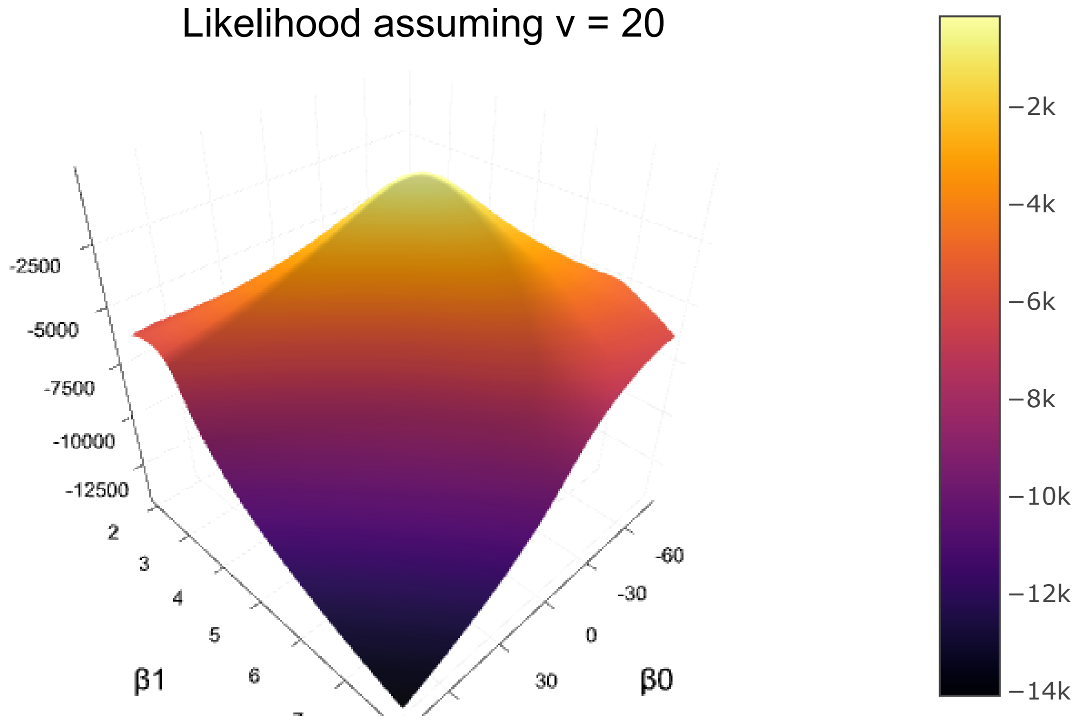
In [4]:
```
scatter(tdata, :x, :y)
```

Out[4]:



In [5]:
```
surface(
    linspace(-80, 80, 100), linspace(2, 8, 100),
    (x, y) -> loglikelihood(x, y, 100, tdata[:x], tdata[:y]),
    xlabel = "β0", ylabel = "β1", title = "Likelihood assuming v = 20"
)
```

Out[5]:




In [6]:
```
using NLopt # API to standard non-linear optimizer
```

In [7]:
```
opt = Opt(:LN_NELDERMEAD, 3)
obj(θ, nograd) = -loglikelihood(θ[1], θ[2], θ[3], tdata[:x], tdata[:y])
lower_bounds!(opt, [-Inf, -Inf, 0.])
min_objective!(opt, obj)
```

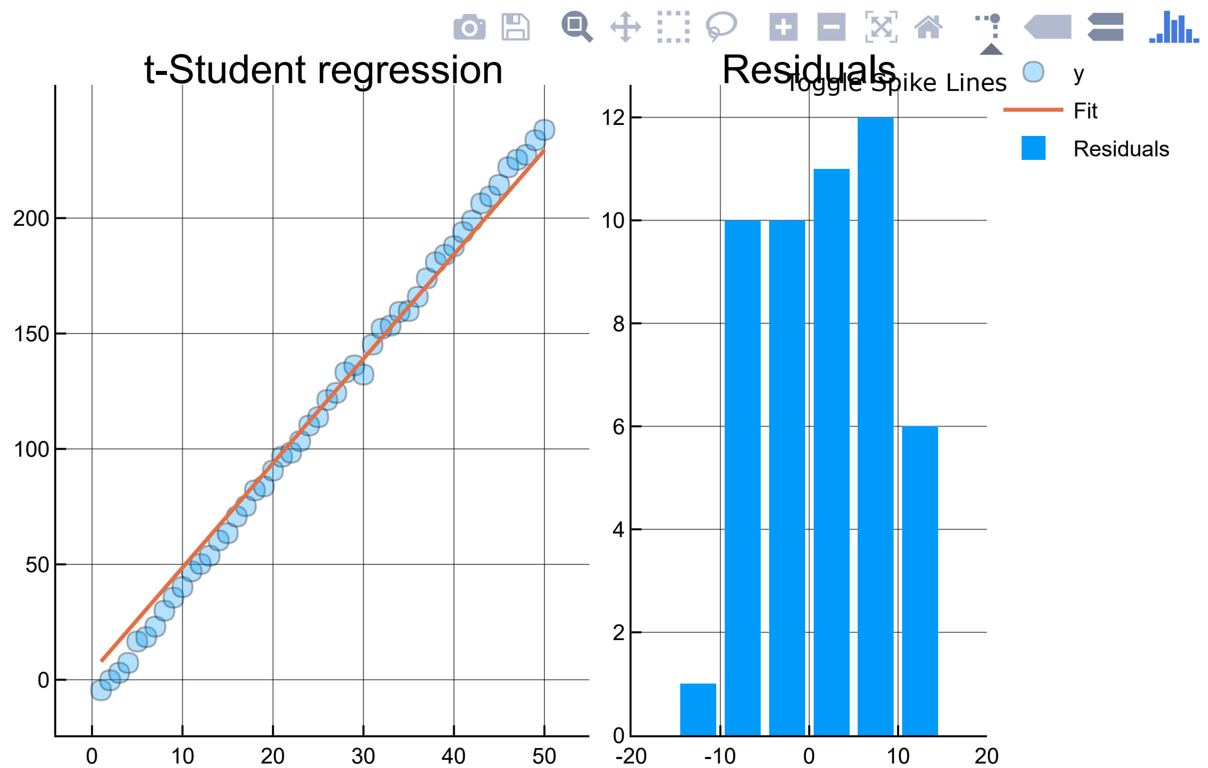
These are the MLE estimates for the t-student fit:

In [8]:
```
tic()
minf, minx, ret = optimize(opt, [1.234, 5.678, 50])
println("MLE Estimates:\n\t β0, β1, v = ", minx)
β0TS, β1TS, v = minx
yTS = β0TS + β1TS*tdata[:x]
toc();
```

```
MLE Estimates:
         β0, β1, v = [3.37292, 4.52193, 0.492522]
elapsed time: 0.889880338 seconds
```

In [9]:
```
resTS = yTS - tdata[:y]
plot(
    plot(tdata[:x], [tdata[:y] yTS],
        seriestype = [:scatter :line], title = "t-Student regression",
        labels = ["y" "Fit"],
        ms = [5 0], alpha = [0.3 1], lw = [1 2]),
    histogram(resTS, title = "Residuals", xlims = (-20,20), labels = "Residuals")
    layout = @layout [a{0.6w} b{0.4w}]
)
```

Out[9]:




We now fit the OLS estimators

In [10]:
```
x = tdata[:x]
y = tdata[:y]
X = [ones(length(x)) x]
β0OLS, β1OLS = Symmetric(X' * X) \ X' * y
```

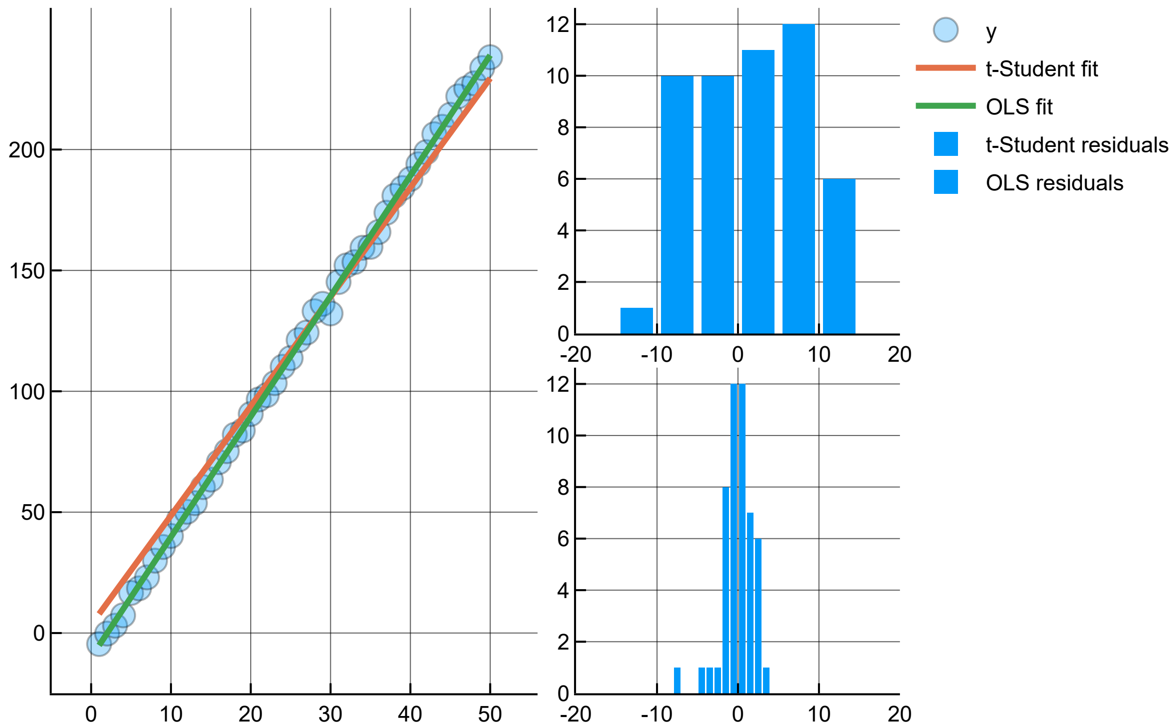
Out[10]:
```
2-element DataArrays.DataArray{Float64,1}:
 -10.0878
   4.98212
```

The largest difference is in the incercept. We can visually compare the two models,

In [11]:

```
yOLS = β0OLS + β1OLS*x
resOLS = y - yOLS
plot(
    plot(x, [y yTS yOLS],
        st = [:scatter :line :line],
        labels = ["y" "t-Student fit" "OLS fit"],
        lw = [1 3 3],
        alpha = [0.3 1 1],
        ms = [6 1 1]
    ),
    histogram(resTS, xlims = (-20,20), labels = "t-Student residuals"),
    histogram(resOLS, xlims = (-20,20), labels = "OLS residuals"),
    layout = @layout [a{0.6w} [b{0.5h}; c{0.5h}]]
)
```

Out[11]:




The OLS weights seem to give a better fit, the t-Student have larger bias at the beginning and end of the curve.