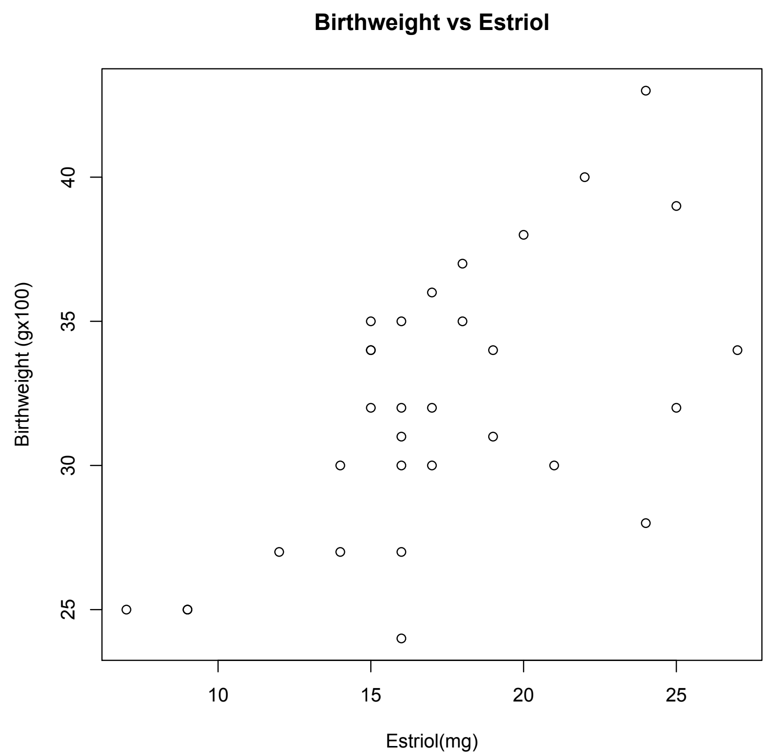
Problem 1:

A compound called estriol was measured over a 24 hour period in the blood of pregnant women. The babies' weights were then recorded at birth. The purpose of the study was to determine whether there was a relationship between estriol and birth weight.[1] The data from this study can be found at http://www.biostat.umn.edu/~lynn/iid/estriol.dat. Estriol was measured in milligrams per 24 hours and birthweight was measured in grams/100.

Using the non-informative prior distribution $g(\eta, \beta, \rho) \propto \rho^{-1}$, find the joint posterior distribution for the transformed intercept η, the slope β, and the precision ρ.

Solution:

A plot of birthweight versus estriol measurement is shown below. Qualitatively, there appears to be an increasing and roughly linear relationship between birthweight and estriol.



The summary statistics for this problem are:

$\overline{X}$	17.23
$\overline{Y}$	32.00
S_{xy}	412.00
S_{xx}	677.42
S_{ee}	423.43

[1] Original source: Greene and Touchstone (1963). 'Urinary tract estriol: an index of placental function,' American Journal of Obstetrics and Gynecology, 85:1-9. Reprinted in Rosner (1982). Fundamentals of biostatistics, Duxbury Press.

Using the formulas from the Unit 8 notes:

- The precision ρ has a Gamma distribution with shape $(n\text{-}2)/2 = 14.5$ and scale $2/S_{ee} = 0.0047$.
- Conditional on ρ, the slope β has a normal distribution with mean $b = S_{xy} / S_{xx} = 0.608$ and precision $S_{xx}\rho = 677.4\rho$.
- Conditional on ρ, the transformed intercept η has a normal distribution with mean $h = \overline{Y} = 32$ and precision $n\rho = 31\rho$.
- Therefore, the joint density function for (η, β, ρ) is the product of a gamma density function for ρ with shape 14.5 and scale 0.0047, a normal density function with mean 32 and precision 31ρ for η, and a normal density function with mean 0.608 and precision 677.4ρ for β.

The distribution of the untransformed intercept $\alpha = \eta - \beta\overline{X}$ given β and ρ is normal with mean $32 - 17.23\beta$ and precision 31ρ.

Some students stated that ρ has a gamma distribution and η and β have nonstandard t distributions. This is true, but these are the marginal distributions not the joint distribution as the problem asked.

***Problem 2*:**

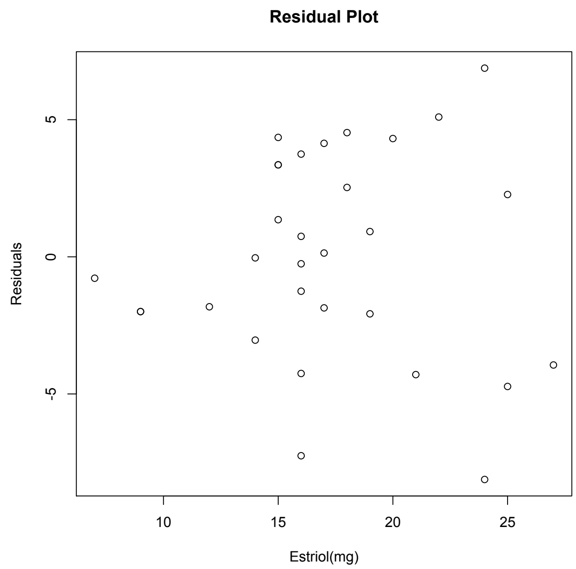
Comment on your results, including whether the assumptions for normal linear regression are met.

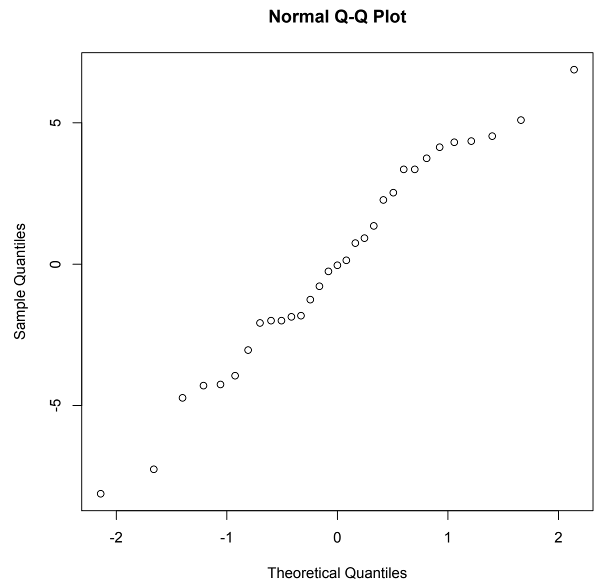
***Solution*:**

A plot of residuals versus estriol measurement is shown below, along with a normal Q-Q plot of the residuals.

The normal Q-Q plot looks fairly linear, and a hypothesis test fails to reject the hypothesis of normal errors.

The residual plot suggests that the error variance may be smaller for smaller concentrations of estriol. Some students did a hypothesis test that rejected the equal variances assumption. On the other hand, the data at the low end is pretty sparse. Although we do have a concern about unequal variances, it seems reasonable at least provisionally to use a normal linear regression with the posterior distribution of the precision, slope and intercept as given above.





***Problem 3*:**

What is the predictive distribution for birthweight given milligrams of estriol? Find a 90% predictive interval for the birthweight of a baby given that 19 mg of estriol were measured in the mother's urine over 24 hours.

***Solution*:**

The posterior predictive distribution for a baby's birthweight given a measurement of x milligrams of estriol from the mother is a nonstandard t distribution with parameters:

- Center: $a + bx = 32 - 17.23\times0.608 + 0.608x = 21.52 + 0.608x$
- Spread: $\sqrt{\left(\frac{(x-17.23)^2}{677.42}+\frac{1}{31}+1\right)\left(\frac{423.43}{29}\right)} = \sqrt{0.022(x-17.23)^2+15.07}$
- Degrees of freedom: 29

For estriol measurement of 19 mg (see Part d), the predictive distribution is nonstandard t with center 33.08, spread 3.89, and degrees of freedom 29.

The theoretical interval can be found by using the quantiles of the t distribution with 29 degrees of freedom: $33.08 + 3.89t_{0.05}$ and $30.08 + 3.89t_{0.95}$. The theoretical interval is [26.5, 39.7]. The interval can be found by using the t quantiles in R:

```
33.08 + qt(c(0.05,0.95),29)*3.89
```

***Problem 4*:**

Use 1000 Monte Carlo samples to find an approximate 90% predictive interval for the birthweight of a baby given that 19 mg of estriol were measured in the mother's urine over 24 hours.

***Solution*:**

We can do this by simulating 1000 realizations of a standard *t* distribution with 29 degrees of freedom, multiplying the result by the spread 3.89, and then adding the center 30.08:

```
pred_19a = 33.08 + 3.89*rt(1000,29)
quantile(pred_19a,c(0.05,0.95))
```

Another way to do this problem is to simulate ρ from a Gamma distribution with shape 14.5 and scale 0.0047; then simulate η from a normal distribution with mean 32 and standard deviation $(32\rho)^{-1/2}$; then simulate β from a normal distribution with mean 0.608 and standard deviation $(677.42\rho)^{-1/2}$; then simulate ε from a normal distribution with mean 0 and standard deviation $\rho^{-1/2}$; and add these together:

```
# Simulate predictive distribution for x=19
rho=rgamma(1000,shape=14.5,scale=0.0047)
eta=rnorm(1000,32,sqrt(1/(31*rho)))
beta=rnorm(1000, 0.608, sqrt(1/(677.42*rho)))
pred_19b=rnorm(1000,eta+beta*(19-17.23),sqrt(1/rho))
quantile(pred_19b,c(0.05,0.95))
```

Results will be different every time, but on one trial the first method gave the interval [26.6, 39.7] and the second method gave the interval [26.7, 39.7]. Both simulated intervals are consistent with the theoretical interval.