Write an R script to solve the problems prefixed with an R. In Word (or any other word processor), answer all the problems prefixed with a W, without using R. Then convert the word processed file (not the R script) into pdf format. Submit both files, named according to the convention given in Homework 0, attached to a single email.

Information on this data set from its source, http://www.statsci.org/data/general/wolfrive.html: “Jaffe, Parker and Wilson have investigated the concentration of several hydrophobic organic substances (such as hexachlorobenzene, chlordane, heptachlor, aldrin, dieldrin, endrin) in the Wolf River in Tennessee. Measurements were taken downstream of an abandoned dump site that had previously been used by the pesticide industry to dispose of its waste products.

It was expected that these hydrophic substances might have a nonhomogeneous vertical distribution in the river because of differences in density between these compounds and water and because of the adsorption of these compounds on sediments, which could lead to higher concentrations on the bottom. It is important to check this hypothesis because the standard procedure of sampling at six-tenths of the depth could miss the bulk of these pollutants if the distribution were not uniform.

Grab samples were taken with a La Motte-Vandorn water sampler of 1 litre capacity at various depths of the river. This sampler consists of a horizontal plexiglas tube of 7 centimetres diameter and a plunger of each side which shuts the sampler when the sampler is at the desired depth. Ten surface, 10 mid-depth and 10 bottom samples were collected, all within a relatively short period. Until they were analysed the samples were stored in 1-quart mason jars at low temperature.

In the analysis of the samples, a 250-millilitre water sample was taken from each mason jar and was extracted with 1 millilitre of either hexanes or petroleum ether. A sample of the extract was then injected into a gas chromatograph and the output was compared against standards of known concentrations. The test procedure was repeated two more times, injecting different samples of the extract in the gas chromatograph. The average aldrin and hexachlorobenzene (HCB) concentrations (in nanograms per liter) in these 30 samples are given in the data.”

We are interested in the question: Does average HCB concentration differ among the three depth levels of the river?

To address this question, we define the random variable $H$ that is the HCB concentration as measured by the process described above, and the categorical variable $D$ whose value is the river depth (Bottom, Middepth, or Surface).

It would be good to think about what an individual is here, what population our statistical inferences extend to, and whether our inferences carry much causative weight.

For reference, here are simultaneous density plots of our observations of $HCB$, grouped by depth level:
Answer the following, but don’t include any R code in your write-up.

W1. With $H$ and $D$ defined as above, how should we denote the true model equation with model formula $H \sim D$ to address our question of interest?

W2. We check that the model fitting assumptions are fine, so we fit a model with model formula $H \sim D$. The model summary for this model is

```
Call:
  lm(formula = HCB ~ Depth, data = riverData)

Residuals:
     Min      1Q  Median      3Q     Max
-1.9490 -0.5072  0.0385  0.6673  1.4400

Coefficients:            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.8390     0.2972  19.647  <2e-16 ***
DepthMiddepth -0.5090     0.4203  -1.211  0.2364
DepthSurface  -1.0350     0.4203  -2.463  0.0205 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9398 on 27 degrees of freedom
Multiple R-squared:  0.1834, Adjusted R-squared:  0.1229
F-statistic: 3.032 on 2 and 27 DF,  p-value: 0.06487
```

What is the model equation for this fitted model?

W3. Based on the fitted model summary, what are our estimates of $\mu[H|S = \text{Bottom}]$, $\mu[H|S = \text{Middepth}]$, and $\mu[H|S = \text{Surface}]$?
W4. A normal quantile plot of the residuals of the fitted model is:

![Normal Quantile Plot](image)

From this plot, does it appear that there are any violations of sampling variability assumptions? Explain how you know. (Discuss only the assumption(s) that can be assessed with this plot.)

W5. Simultaneous density plots of the standardized residuals grouped by depth level are:

![Density Plots](image)

From this plot, does it appear that there are any violations of sampling variability assumptions? Explain how you know. (Discuss only the assumption(s) that can be assessed with this plot.)

W6. Assuming that the sampling variability assumptions don’t appear to be violated (which may or may not be the case — see the previous questions), report and interpret the test(s)
conducted as part of the model summary that addresses our main question of interest most closely. In answering this question, as usual state the null hypothesis of each test (but you don’t need to include the alternative hypothesis), and include the value of the test statistic, the degrees of freedom, the \( p \)-value, and what this did or did not constitute statistically significant evidence against. Your answer should be statistical in its language, using terms such as \( \mu[H|D = \text{Surface}] \), for example.