One-Way ANOVA Diagnostics

Allison Theobold

Assumptions of a One-Way ANOVA

1. Independence of observations
2. Variances (SD) between groups are the same
3. (F-test) Normality of responses (or errors)

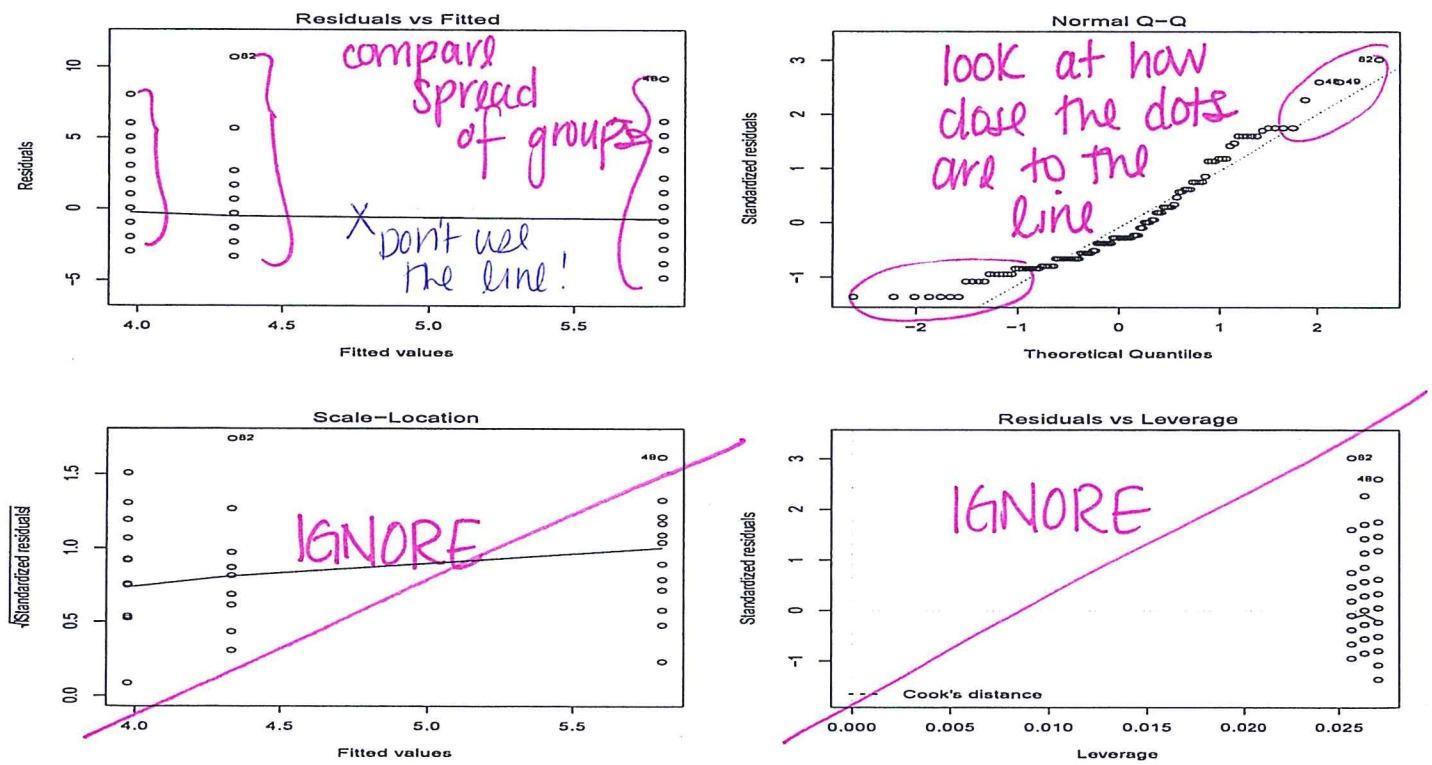
Balanced Design:

Every group has the same sample size.

* Not an assumption, but can influence your test results!

R's built-in diagnostic plots:

```
require(heplots)
data(MockJury)
mockJury.model <- lm(Years ~ Attr, data = MockJury)
par(mfrow = c(2, 2))
plot(mockJury.model)
```



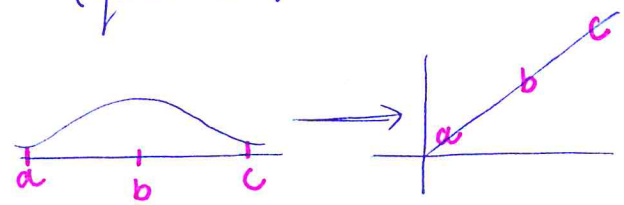
- **Residuals vs Fitted Plot**

Displays residuals on y-axis & fitted values on x-axis. Look for: equal spread of dots between groups!

- **Normal Q-Q Plot**

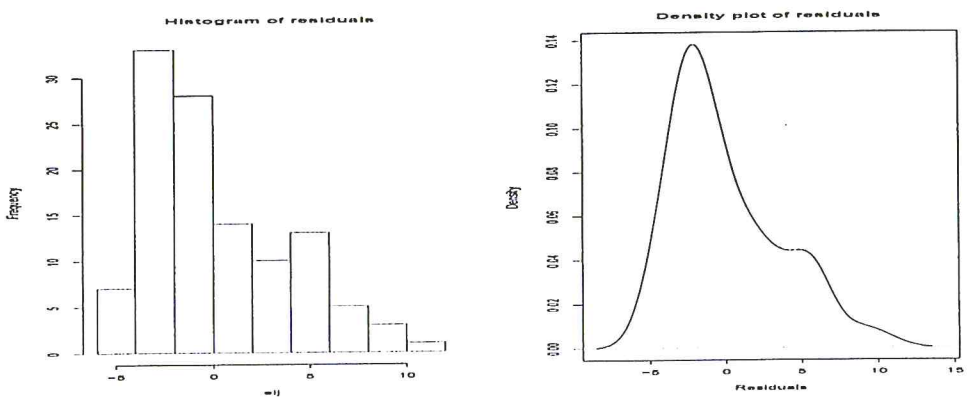
Displays errors (residuals) on a Normal distribution (quantiles)

Look for: points following the line fairly close

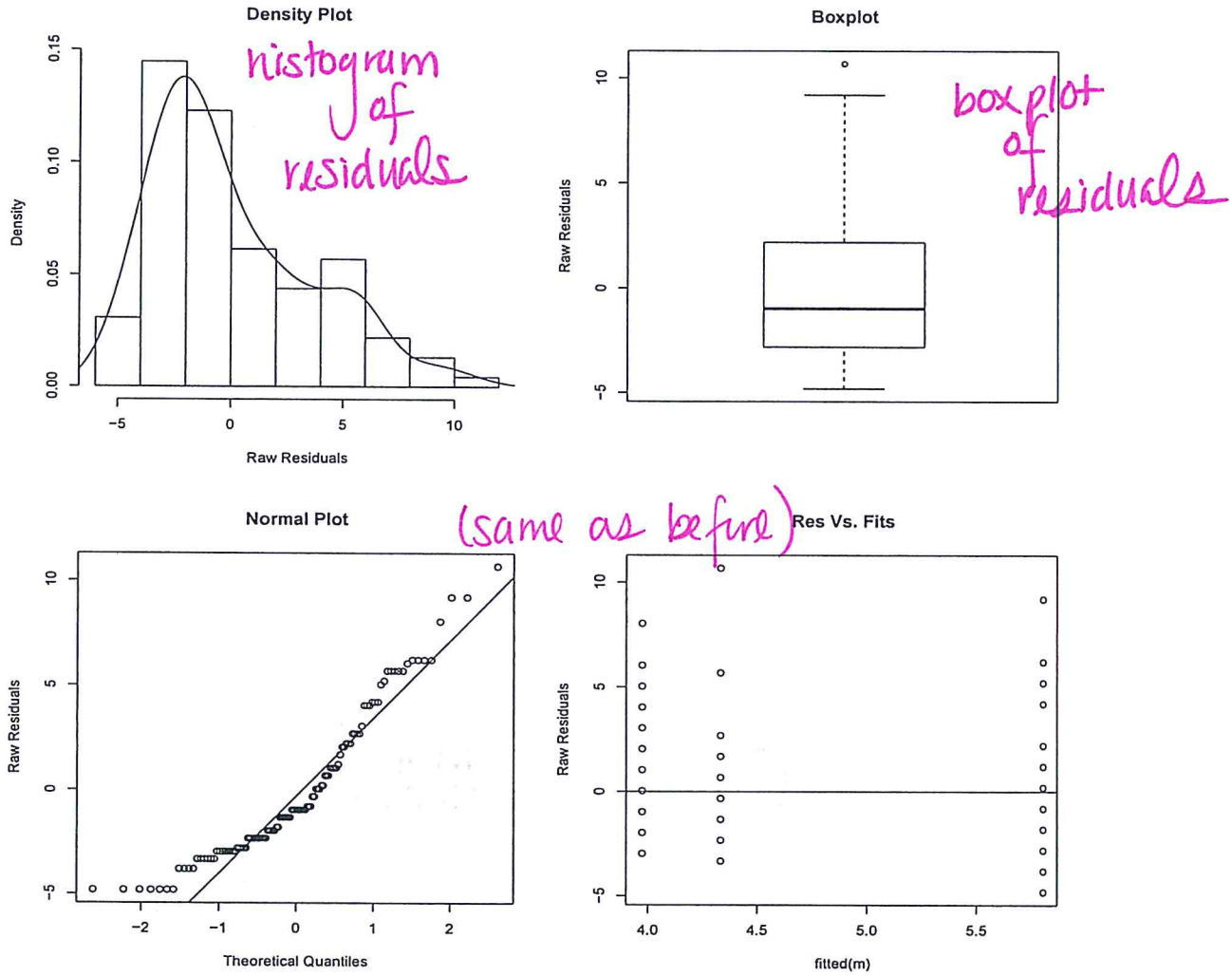


Another way to look at the distribution of the residuals:

```
par(mfrow = c(1, 2))
eij <- residuals(mockJury.model)
hist(eij, main = "Histogram of residuals")
plot(density(eij), main = "Density plot of residuals", ylab = "Density",
    xlab = "Residuals")
```



Al's diagANOVA Function



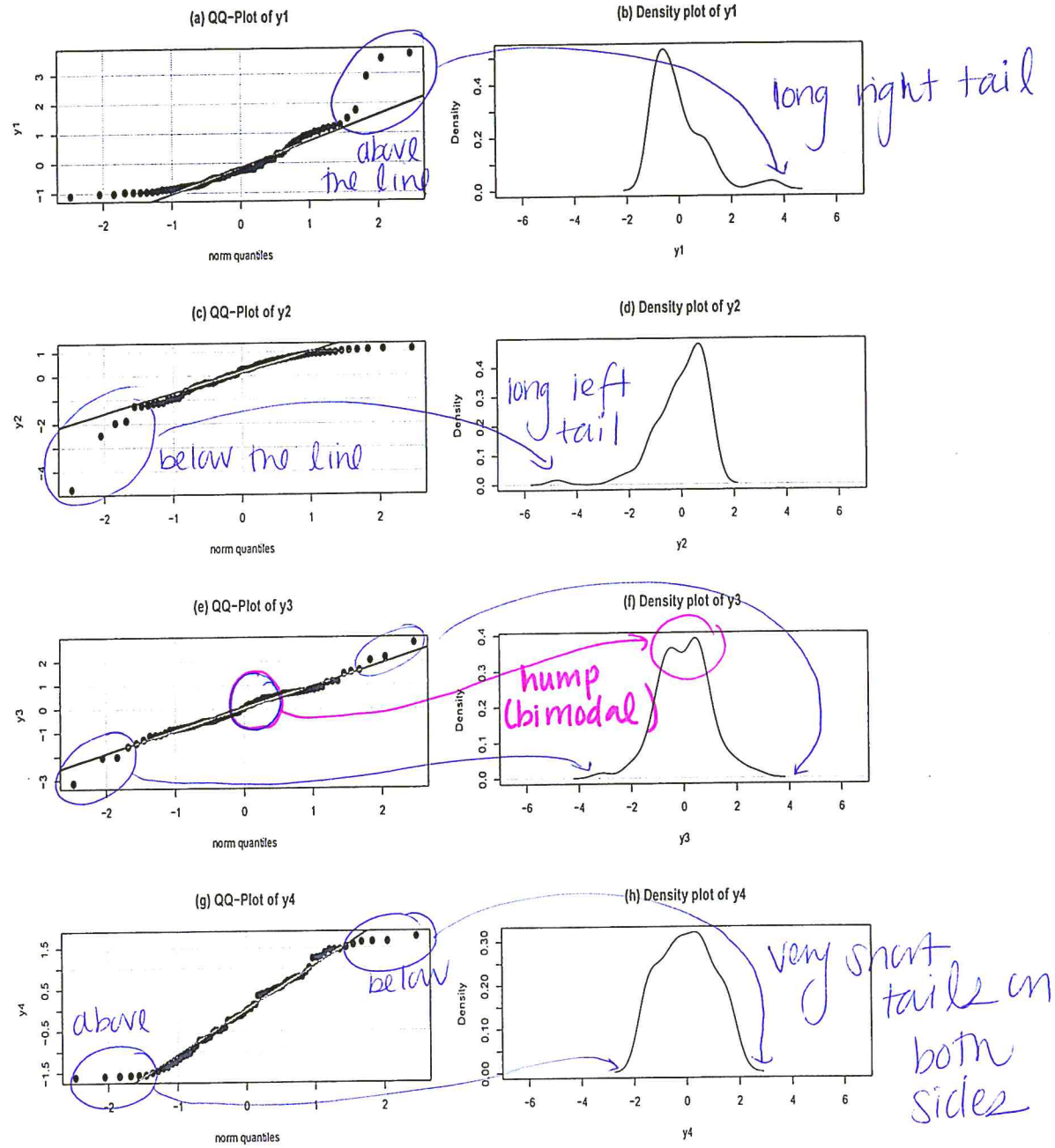

```
## [1] "In this sample of size n=114, correlation of the residuals in the qq-plot is r=0.957673"
## [1] "In the following table, if r < critical.r, then the qq-plot suggests the residuals are not normal:"
##     n critical.r
## 1   5      0.832
## 2  10      0.880
## 3  15      0.911
## 4  20      0.929
## 5  25      0.941
## 6  30      0.949
## 7  40      0.960
## 8  50      0.966
## 9  60      0.971
## 10 75      0.976
##
##  Shapiro-Wilk normality test
##
## data:  r
## W = 0.9, p-value = 2e-06
```

correlation check for Normality

If you use diagANOVA with do.tests = TRUE you get these!

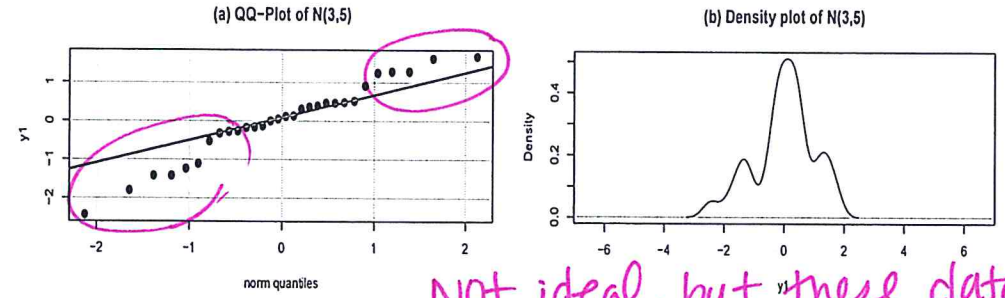
you won't use this.

Some examples of Q-Q plots:

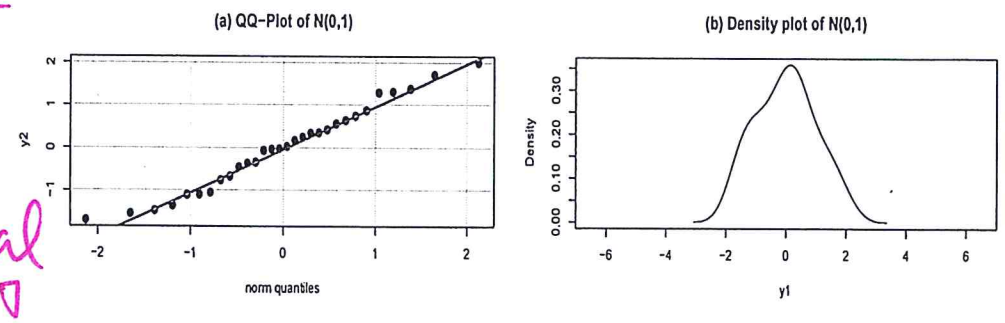


Some examples of Normal Q-Q plots:

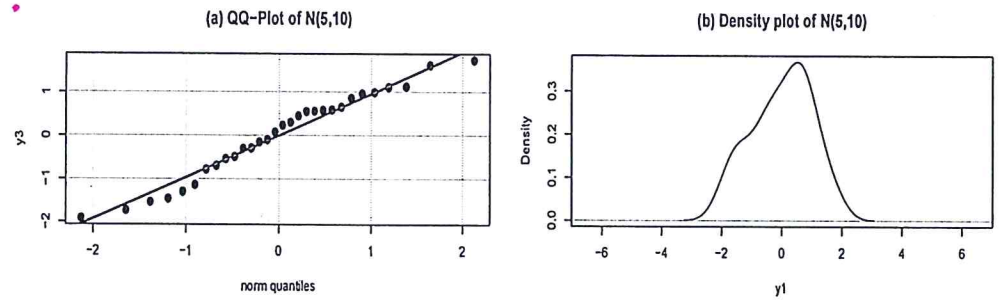
These are Q-Q plots that come from Normal data!

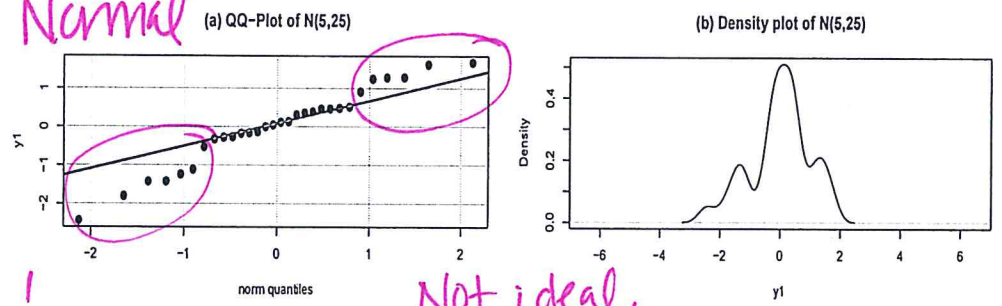


Not ideal, but these data are Normal



Keep in mind even Normal data can look funky!





Not ideal, but these data are Normal!

Checking the balance of the design:

```
table(MockJury$Attr)
```

← table will get you counts in each group.

```
##
##    Beautiful      Average Unattractive
##           39           38           37
```

Unequal sample sizes = unbalanced design