Assignment #1

Jeremy Yeaton

November 25, 2017

Question 1:

(a) *Give a numerical summary of FEV1 (mean, standard deviation and range) for each smoking category (recoded as a categorical variable with appropriate levels), and for all subjects (grand mean and overall standard deviation). Results should be printed in one or two Tables.*

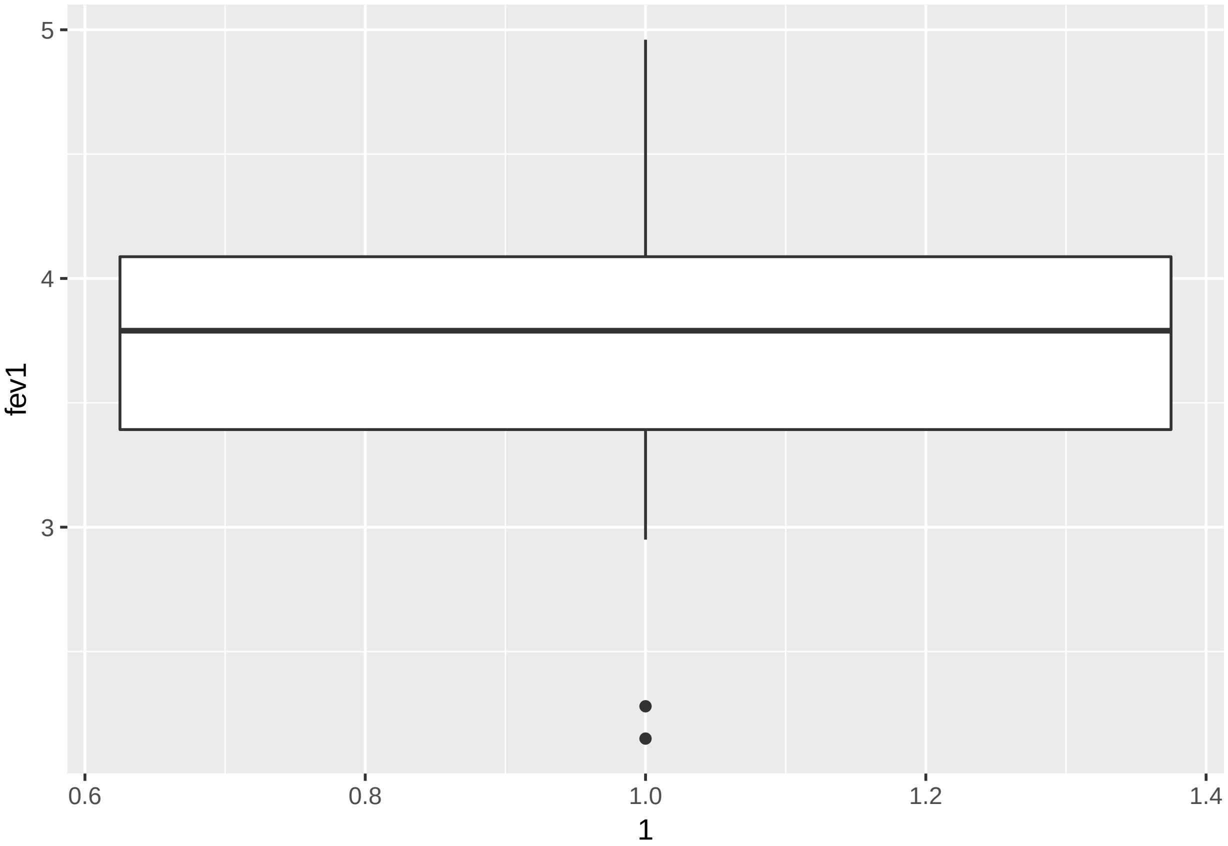
The table below shows the descriptive statistics for each of the groups, as well as a summary for the whole sample directly below.

cat.f	mean	sd	range
current	3.220000	0.6758106	1.82
early	3.938333	0.2545912	0.69
non-smoker	4.220000	0.5726081	1.46
recent	3.460000	0.7128534	2.12

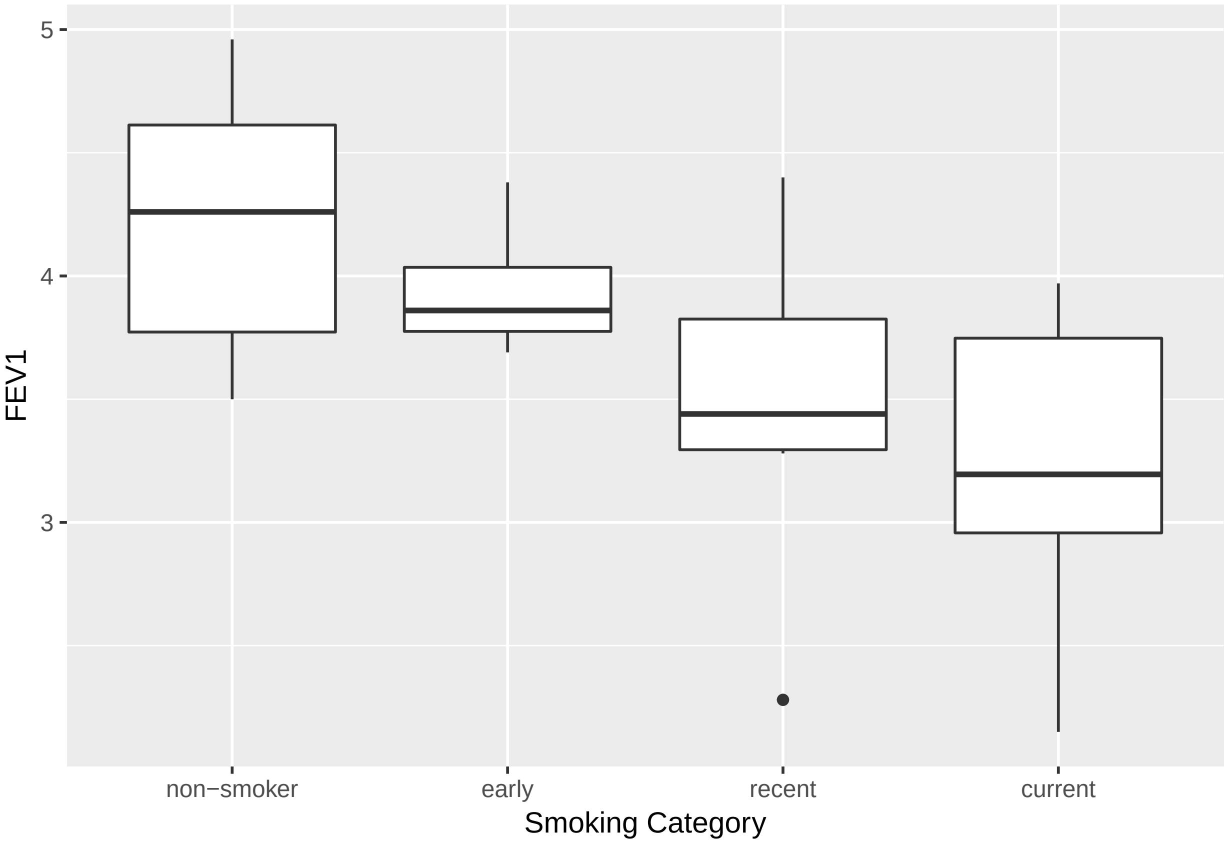
grand mean	overall sd	total range
3.709583	0.6749202	2.81

(b) *Use box-and-whiskers charts or density plots to show the distribution of individual values.*

Below is a box-and-whisker plot of all of the points in the sample.



The following box-and-whisker chart displays the distribution of the individual values by group, and provides a more useful insight into the data than the one above.



Question 2:

Carry out a one-way ANOVA to test the null hypothesis that FEV1 does not depend on smoking category.

```
by_cat.anova
```

```
## $ANOVA
##   Effect DFn DFd        F          p p<.05       ges
## 1  cat.f   3  20 3.623135 0.03085325     * 0.3521093
##
## $`Levene's Test for Homogeneity of Variance`
##   DFn DFd       SSn      SSd        F         p p<.05
## 1   3  20 0.4522792 2.229833 1.352206 0.2859163
```

(a) *Formulate your conclusion in plain English, and*

Based on our ANOVA, it seems to be the case that FEV1 *does* depend on smoking category, with a very low p-value and a high F-value, and as such, we can reject the null-hypothesis.

(b) *report the percentage of explained variance.*

The percentage of variance explained by this test is about 35.2%.

Question 3:

(a) *Use post-hoc Tukey HSD tests (R command: TukeyHSD) to compare all pairs of means among the four groups of smokers.*

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## cat.f        3  3.689  1.2297   3.623 0.0309 *
## Residuals   20  6.788  0.3394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = fev1 ~ cat.f, data = f_clean)
##
## $cat.f
##                          diff         lwr       upr     p adj
## early-current       0.7183333 -0.22308912 1.6597558 0.1760748
## non-smoker-current  1.0000000  0.05857755 1.9414224 0.0348503
## recent-current      0.2400000 -0.70142245 1.1814224 0.8905477
## non-smoker-early    0.2816667 -0.65975578 1.2230891 0.8360677
## recent-early       -0.4783333 -1.41975578 0.4630891 0.5008038
## recent-non-smoker  -0.7600000 -1.70142245 0.1814224 0.1415657
```

```
## Warning in plot.window(...): "fig.height" is not a graphical parameter
```

```
## Warning in plot.window(...): "fig.width" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "fig.height" is not a graphical
## parameter
```

```
## Warning in plot.xy(xy, type, ...): "fig.width" is not a graphical parameter
```

```
## Warning in title(...): "fig.height" is not a graphical parameter
```

```
## Warning in title(...): "fig.width" is not a graphical parameter
```

```
## Warning in axis(1, ...): "fig.height" is not a graphical parameter
```

```
## Warning in axis(1, ...): "fig.width" is not a graphical parameter
```

```
## Warning in axis(2, at = nrow(xi):1, labels = dimnames(xi)[[1L]], srt = 0, :
## "fig.height" is not a graphical parameter
```

```
## Warning in axis(2, at = nrow(xi):1, labels = dimnames(xi)[[1L]], srt = 0, :
## "fig.width" is not a graphical parameter
```

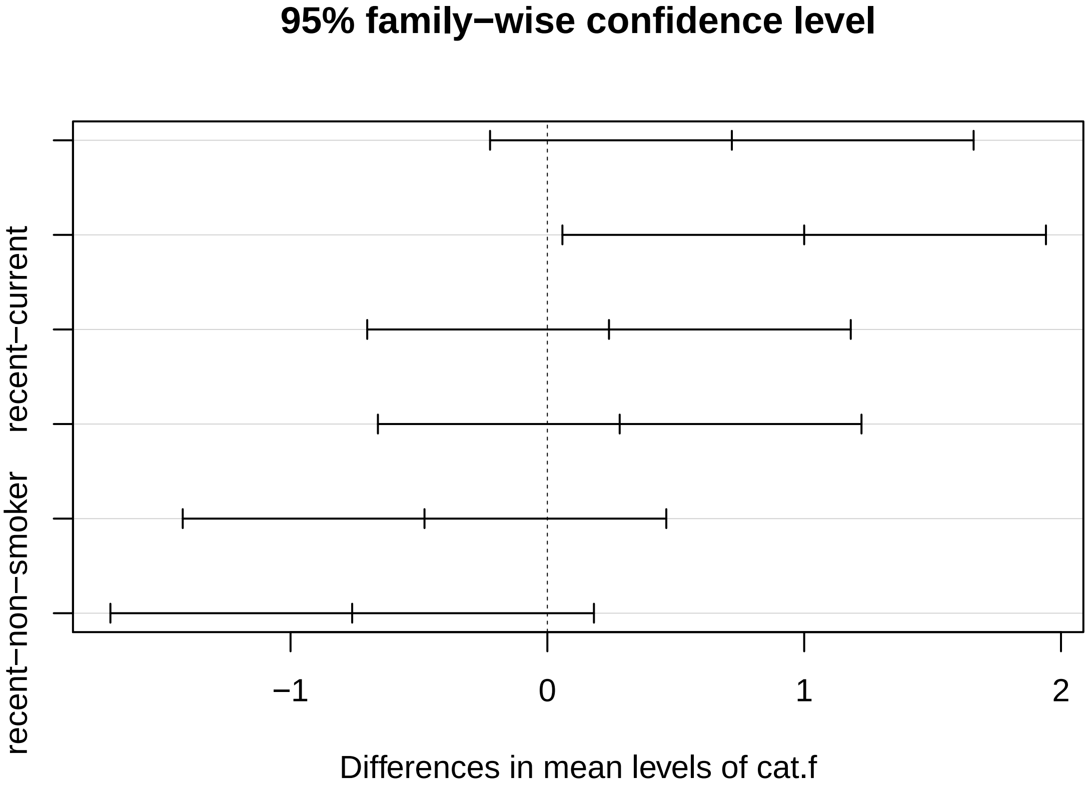
```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "fig.height" is not a graphical parameter
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "fig.width" is not a graphical parameter
```

```
## Warning in segments(xi[, "lwr"], yvals, xi[, "upr"], yvals, ...):
## "fig.height" is not a graphical parameter
```

```
## Warning in segments(xi[, "lwr"], yvals, xi[, "upr"], yvals, ...):
## "fig.width" is not a graphical parameter
```

```
## Warning in segments(as.vector(xi), rep.int(yvals - 0.1, 3L),
## as.vector(xi), : "fig.height" is not a graphical parameter
## Warning in segments(as.vector(xi), rep.int(yvals - 0.1, 3L),
## as.vector(xi), : "fig.width" is not a graphical parameter
```



Summarize point estimates and 95% confidence intervals in a Table or graphical display,and indicate which pairs of means are found to be significantly different.

(b) *Compare those results with results from all pairwise comparisons for mean FEV1 using the Bonferroni method (R command: pairwise.t.test).*

Question 4:

Is there any evidence for a linear or quadratic trend for mean FEV1 when considering smoking status as ordered factor levels: 1 < 2 < 3 < 4 (use the R command factor with the ordered = TRUE option

Question 5:

(a) *Compare the preceding results with the conclusion that would be reached by using a regression approach where one considers smoking status as a numerical variable, as well as its square, i.e., using the R command lm with a formula like FEV1 ~ smoking + I(smoking)ˆ2.*

(b) *What could explain the difference, if any?*

Appendix

```
f_clean
```

```
##    cat fev1      cat.f ident
## 1    1 4.41 non-smoker     1
## 2    1 4.96 non-smoker     2
## 3    1 3.50 non-smoker     3
## 4    1 3.66 non-smoker     4
## 5    1 4.68 non-smoker     5
## 6    1 4.11 non-smoker     6
## 7    2 3.69      early     7
## 8    2 3.90      early     8
## 9    2 3.82      early     9
## 10   2 4.08      early    10
## 11   2 3.76      early    11
## 12   2 4.38      early    12
## 13   3 3.54     recent    13
## 14   3 4.40     recent    14
## 15   3 3.28     recent    15
## 16   3 2.28     recent    16
## 17   3 3.34     recent    17
## 18   3 3.92     recent    18
## 19   4 2.98    current    19
## 20   4 2.95    current    20
## 21   4 2.15    current    21
## 22   4 3.41    current    22
## 23   4 3.97    current    23
## 24   4 3.86    current    24
```