

Assignment 1 Solutions

1. Question 1

(a) Initial plotting

Figure 1: Histograms of Tornado Death

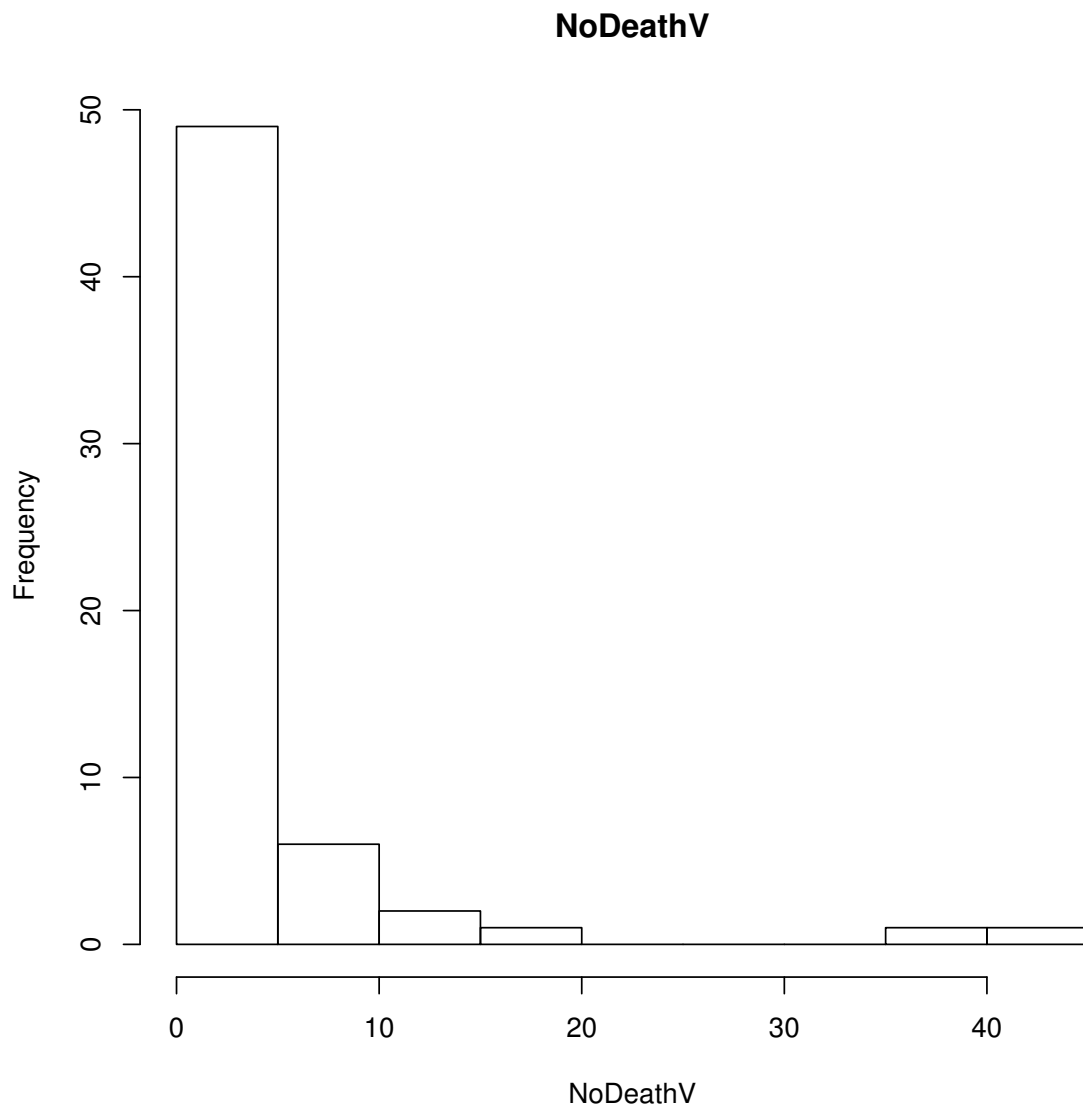
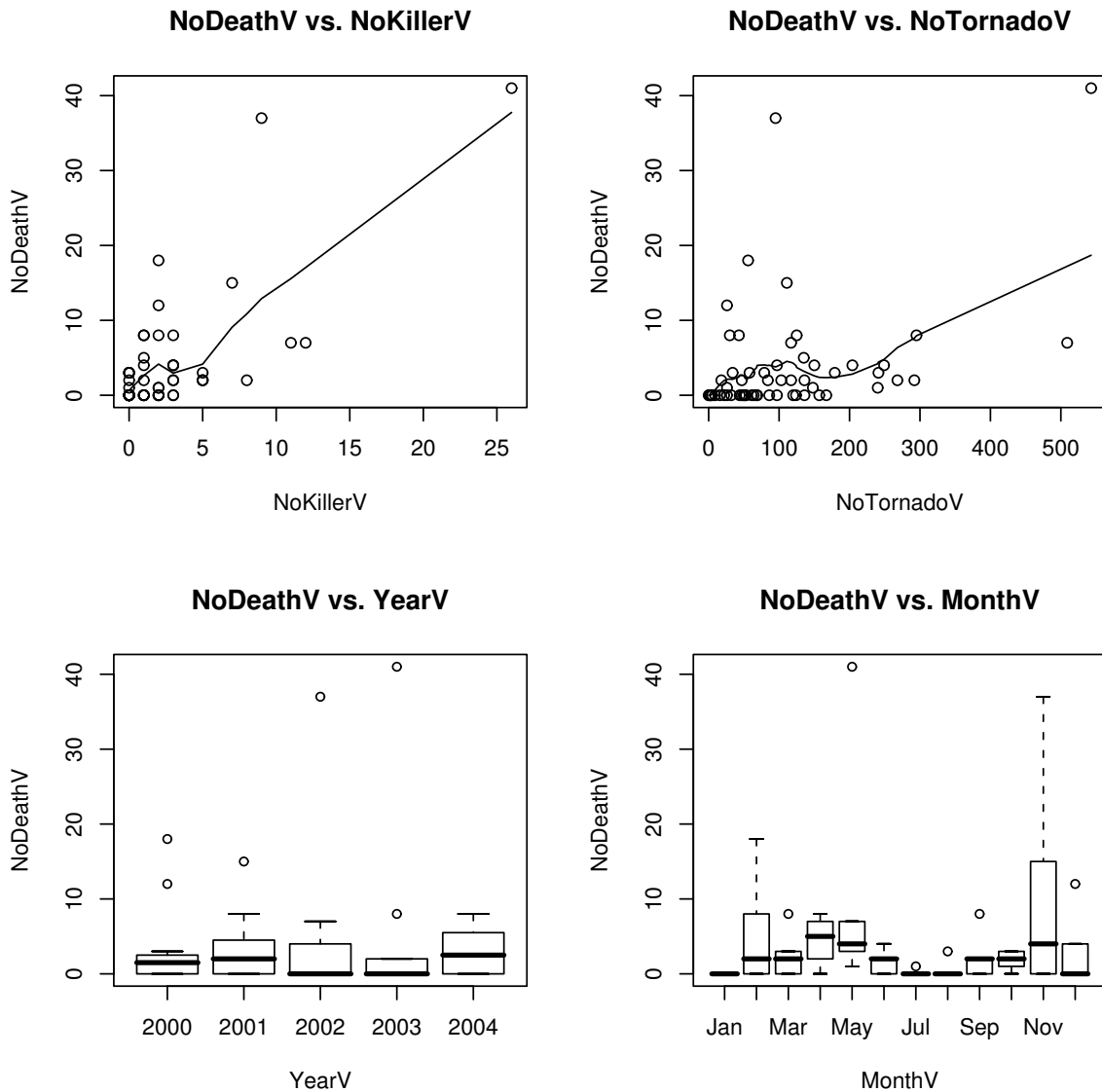


Figure 2: Preliminary Examination



Additive model similar to page 39 and 41

The F-test results are very similar to those of the model in Page 39 and 41. Based on the F-test results, only the variable Number of Killer Tornadoes is significant. The results for F-test, as well as the point estimates and confidence intervals of the regression parameters are listed below using output from R.

```
lm(formula = NoDeathV ~ NoKillerV + NoTornadoV + YearV + MonthV)
```

Analysis of Variance Table

Response: NoDeathV

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NoKillerV	1	1854.80	1854.80	65.6351	4.039e-10 ***
NoTornadoV	1	83.75	83.75	2.9636	0.09252 .
YearV	4	1.98	0.49	0.0175	0.99938
MonthV	11	302.33	27.48	0.9726	0.48531
Residuals	42	1186.89	28.26		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residuals:

Min	1Q	Median	3Q	Max
-8.7869	-2.3502	-0.1213	2.0383	17.7531

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.422861	2.754429	-0.154	0.8787
NoKillerV	1.368921	0.292717	4.677	3.02e-05 ***
NoTornadoV	0.009503	0.017066	0.557	0.5806
YearV2001	-0.682039	2.175108	-0.314	0.7554
YearV2002	0.448250	2.200312	0.204	0.8396
YearV2003	-0.440610	2.192109	-0.201	0.8417
YearV2004	-1.574626	2.379108	-0.662	0.5117
MonthVFeb	4.885186	3.370300	1.449	0.1546
MonthVMar	1.029734	3.436574	0.300	0.7659
MonthVApr	-1.476362	3.753309	-0.393	0.6961
MonthVMay	-4.370168	5.570040	-0.785	0.4371
MonthVJun	-3.614433	4.520009	-0.800	0.4284
MonthVJul	-2.035446	3.858593	-0.528	0.6006
MonthVAug	0.381660	3.665029	0.104	0.9176
MonthVSep	0.096025	3.678179	0.026	0.9793
MonthVOct	-0.171384	3.481616	-0.049	0.9610
MonthVNov	5.998431	3.527514	1.700	0.0964 .
MonthVDec	2.650641	3.392925	0.781	0.4391

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Confidence Intervals:

	2.5 %	97.5 %
(Intercept)	-5.98152371	5.13580148
NoKillerV	0.77819329	1.95964809
NoTornadoV	-0.02493849	0.04394365
YearV2001	-5.07158478	3.70750692
YearV2002	-3.99216042	4.88865969

YearV2003	-4.86446470	3.98324392
YearV2004	-6.37586058	3.22660768
MonthVFeb	-1.91635357	11.68672623
MonthVMar	-5.90555284	7.96502179
MonthVApr	-9.05084615	6.09812203
MonthVMay	-15.61096359	6.87062684
MonthVJun	-12.73618056	5.50731400
MonthVJul	-9.82240196	5.75150979
MonthVAug	-7.01466813	7.77798888
MonthVSep	-7.32684099	7.51889198
MonthVOct	-7.19757036	6.85480190
MonthVNov	-1.12038045	13.11724231
MonthVDec	-4.19655858	9.49784070

(b) **Model after combining non-tornado months**

The modelling results when the non-tornado months (July to December) are combined as a category 'Jul-Dec'. The results for F-test, as well as the point estimates and confidence intervals of the regression parameters are listed below using output from R.

```
lm(formula = NoDeathV ~ NoKillerV + NoTornadoV + YearV + MonthV1)
```

Analysis of Variance Table

Response: NoDeathV

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NoKillerV	1	1854.80	1854.80	64.0804	2.504e-10 ***
NoTornadoV	1	83.75	83.75	2.8934	0.09555 .
YearV	4	1.98	0.49	0.0171	0.99941
MonthV1	6	128.81	21.47	0.7417	0.61883
Residuals	47	1360.41	28.94		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residuals:

Min	1Q	Median	3Q	Max
-11.122	-2.198	-0.835	2.028	21.436

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.594212	2.785756	-0.213	0.832
NoKillerV	1.595446	0.274018	5.822	4.99e-07 ***
NoTornadoV	-0.002889	0.015464	-0.187	0.853

YearV2001	-0.630824	2.200713	-0.287	0.776
YearV2002	0.152665	2.221042	0.069	0.945
YearV2003	-0.356315	2.216097	-0.161	0.873
YearV2004	-0.902803	2.368352	-0.381	0.705
MonthV1Feb	5.012667	3.409563	1.470	0.148
MonthV1Mar	1.468090	3.464505	0.424	0.674
MonthV1Apr	-0.652999	3.739625	-0.175	0.862
MonthV1May	-2.170996	5.350032	-0.406	0.687
MonthV1Jun	-1.642615	4.378670	-0.375	0.709
MonthV1Jul-Dec	1.921125	2.813370	0.683	0.498

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.38 on 47 degrees of freedom

Multiple R-Squared: 0.6033, Adjusted R-squared: 0.5021

F-statistic: 5.958 on 12 and 47 DF, p-value: 3.629e-06

Confidence Intervals:

	2.5 %	97.5 %
(Intercept)	-6.19842920	5.010005
NoKillerV	1.04419375	2.146698
NoTornadoV	-0.03399789	0.028220
YearV2001	-5.05808729	3.796439
YearV2002	-4.31549569	4.620826
YearV2003	-4.81452774	4.101898
YearV2004	-5.66731201	3.861706
MonthV1Feb	-1.84648825	11.871822
MonthV1Mar	-5.50159486	8.437775
MonthV1Apr	-8.17615333	6.870156
MonthV1May	-12.93387187	8.591880
MonthV1Jun	-10.45136255	7.166132
MonthV1Jul-Dec	-3.73864540	7.580895

Hypothesis Testing

Denote $(\beta_1, \beta_2, \dots, \beta_{12})$ as the coefficients for months 1-12 in the first model, then the testing hypothesis can be specified as:

$$H_0 : \beta_7 = \beta_8 = \dots = \beta_{12}$$

$$H_1 : \text{Not all of the } \beta_7 - \beta_{12} \text{ are equal}$$

As the *p-value* of the F-test is 0.3130, we can not reject the null hypothesis. The analysis of variance for the two models in a) and b) are given below.

Analysis of Variance Table

Model 1: NoDeathV ~ NoKillerV + NoTornadoV + YearV + MonthV
 Model 2: NoDeathV ~ NoKillerV + NoTornadoV + YearV + MonthV1

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	42	1186.89				
2	47	1360.41	-5	-173.52	1.2281	0.3130

(c) AIC table

The following three models similar to the ones for page 48 are fitted in R.

```
lm(formula = NoDeathV ~ NoKillerV)
lm(formula = NoDeathV ~ NoKillerV + MonthV1)
lm(formula = NoDeathV ~ NoKillerV * MonthV1)
```

R gives only the *AIC* and the corresponding degrees of freedom for the above three models.

	df	AIC
lm(NoDeathV ~ NoKillerV)	3	372.3302
lm(NoDeathV ~ NoKillerV + MonthV1)	9	376.1348
lm(NoDeathV ~ NoKillerV * MonthV1)	15	372.1501

We can use the formula (3.13) in the textbook to calculate the AIC_c values. The formula is

$$AIC_c = AIC + \frac{2\nu(\nu + 1)}{n - \nu - 1},$$

where for our models in consideration, $n = 60$ and $\nu = p + 2$ equals the degree of freedom given in the above R output for each model. Plugging in the values, we can obtain the complete *AIC* and AIC_c table as follows.

	df	AIC	AICc
lm(NoDeathV ~ NoKillerV)	3	372.3302	372.7588
lm(NoDeathV ~ NoKillerV + MonthV1)	9	376.1348	379.7348
lm(NoDeathV ~ NoKillerV * MonthV1)	15	372.1501	383.0592

It is then easy to get the *AIC* difference table similar to the last three lines of the table in 3.2(c).

	AIC difference	AICc difference
lm(NoDeathV ~ NoKillerV)	0.1801	0
lm(NoDeathV ~ NoKillerV + MonthV1)	3.9847	6.9760
lm(NoDeathV ~ NoKillerV * MonthV1)	0	10.3004

For these models, *AIC* and AIC_c give contradictory results as the last model is complicated with relatively large number of parameters (i.e., 15 degree of freedom).