Anova Exercise
Lincoln Nordquist

Exercise problem 1:

Question 1:

The Anova test I performed gave me an output of **(9.143244123864966 0.00011576580752474306)**, indicating that there were differences in the means between countries. I then performed a Tukey-Kramer post hoc test to determine exactly which pair(s) of countries were differing from each other. Here is the output of my test:

```
==================================================
group1   group2   meandiff p-adj   lower    upper    reject
--------------------------------------------------
Canada France  -0.0724 0.3845 -0.2012  0.0564  False
Canada    USA  -0.1699 0.0004 -0.2734 -0.0664   True
France    USA  -0.0975 0.0407 -0.1917 -0.0032   True
--------------------------------------------------
```

Based on my equation, we can determine that there was a significant rating difference between Canada and USA, France and USA, but NOT Canada and France.

Question 2:

For this question, I needed to perform an Anova test on not only three groups, but any possible number of years, so I tested every possible year combination. Here is my output:

```
ANOVA p-value: 0.00020492752425445773
Multiple Comparison of Means - Tukey HSD, FWER=0.05
==================================================
group1 group2 meandiff p-adj   lower  upper  reject
--------------------------------------------------
  2006   2007   0.0373    1.0 -0.2174 0.2921  False
  2006   2008  -0.1304 0.8448 -0.3743 0.1135  False
  2006   2009  -0.0518 0.9999 -0.2824 0.1787  False
  2006   2010   0.0236    1.0 -0.2115 0.2588  False
  2006   2011   0.1311 0.7244 -0.0884 0.3505  False
  2006   2012   0.0532 0.9997 -0.1611 0.2675  False
  2006   2013    0.072 0.9951  -0.144  0.288  False
  2006   2014   0.0643 0.9975 -0.1438 0.2724  False
  2006   2015   0.1215 0.7338 -0.0834 0.3264  False
  2006   2016    0.101  0.921   -0.11 0.3121  False
```

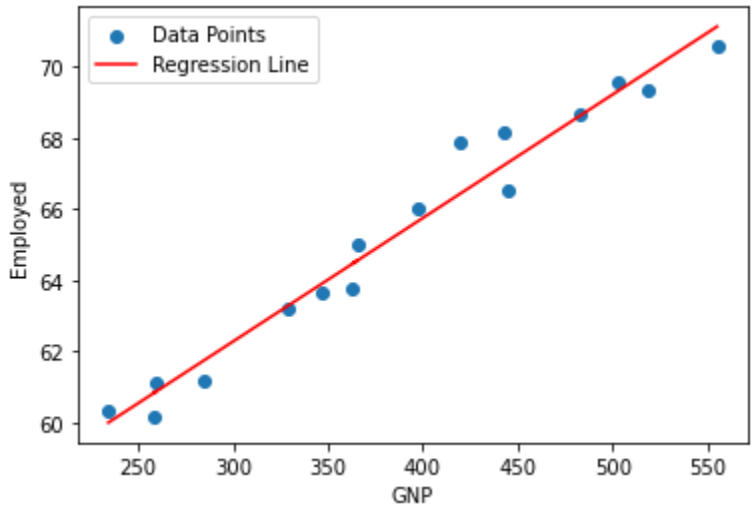
2006	2017	0.1875	0.8794	-0.1787	0.5537	False
2007	2008	-0.1677	0.4822	-0.4071	0.0717	False
2007	2009	-0.0892	0.9801	-0.3149	0.1366	False
2007	2010	-0.0137	1.0	-0.2441	0.2167	False
2007	2011	0.0937	0.9576	-0.1207	0.3082	False
2007	2012	0.0159	1.0	-0.1933	0.225	False
2007	2013	0.0347	1.0	-0.1762	0.2456	False
2007	2014	0.0269	1.0	-0.1759	0.2297	False
2007	2015	0.0842	0.9673	-0.1154	0.2837	False
2007	2016	0.0637	0.9974	-0.1422	0.2695	False
2007	2017	0.1502	0.9718	-0.2131	0.5134	False
2008	2009	0.0785	0.9887	-0.135	0.2921	False
2008	2010	0.154	0.4714	-0.0644	0.3724	False
2008	2011	0.2614	0.0014	0.06	0.4629	True
2008	2012	0.1836	0.0908	-0.0122	0.3794	False
2008	2013	0.2024	0.0392	0.0047	0.4001	True
2008	2014	0.1946	0.0369	0.0056	0.3837	True
2008	2015	0.2519	0.0006	0.0663	0.4374	True
2008	2016	0.2314	0.0049	0.0391	0.4237	True
2008	2017	0.3179	0.1332	-0.0379	0.6736	False
2009	2010	0.0755	0.9879	-0.1279	0.2789	False
2009	2011	0.1829	0.0563	-0.0022	0.368	False
2009	2012	0.105	0.7459	-0.0739	0.2839	False
2009	2013	0.1238	0.5209	-0.0571	0.3048	False
2009	2014	0.1161	0.5382	-0.0554	0.2876	False
2009	2015	0.1733	0.0352	0.0057	0.3409	True
2009	2016	0.1529	0.1575	-0.0222	0.3279	False
2009	2017	0.2393	0.5067	-0.1074	0.586	False
2010	2011	0.1074	0.794	-0.0833	0.2981	False
2010	2012	0.0296	1.0	-0.1552	0.2143	False
2010	2013	0.0484	0.9995	-0.1384	0.2351	False
2010	2014	0.0406	0.9999	-0.1369	0.2182	False
2010	2015	0.0978	0.7946	-0.076	0.2717	False
2010	2016	0.0774	0.9639	-0.1036	0.2584	False
2010	2017	0.1639	0.9314	-0.1859	0.5136	False
2011	2012	-0.0779	0.9261	-0.2422	0.0865	False
2011	2013	-0.059	0.9917	-0.2256	0.1075	False
2011	2014	-0.0668	0.9639	-0.223	0.0894	False
2011	2015	-0.0096	1.0	-0.1616	0.1424	False
2011	2016	-0.03	1.0	-0.1902	0.1301	False
2011	2017	0.0564	1.0	-0.283	0.3959	False
2012	2013	0.0188	1.0	-0.1409	0.1785	False
2012	2014	0.0111	1.0	-0.1378	0.1599	False
2012	2015	0.0683	0.927	-0.0761	0.2127	False

2012	2016	0.0478	0.9972	-0.1052	0.2008	False
2012	2017	0.1343	0.9783	-0.2018	0.4704	False
2013	2014	-0.0077	1.0	-0.159	0.1436	False
2013	2015	0.0495	0.9946	-0.0975	0.1964	False
2013	2016	0.029	1.0	-0.1264	0.1844	False
2013	2017	0.1155	0.9938	-0.2217	0.4527	False
2014	2015	0.0572	0.9662	-0.0778	0.1923	False
2014	2016	0.0368	0.9996	-0.1075	0.181	False
2014	2017	0.1232	0.9879	-0.209	0.4554	False
2015	2016	-0.0205	1.0	-0.1601	0.1192	False
2015	2017	0.066	1.0	-0.2642	0.3962	False
2016	2017	0.0865	0.9995	-0.2476	0.4206	False

As we can see, the ANOVA p-value is 0.00020492752425445773, indicating that there was significant differences in means between multiple groups. The chart I have provided tells which years did and didn't have significant differences. The year combinations that did NOT have significant differences were:

2009 and 2015
2008 and 2013
2008 and 2014
2008 and 2015
2008 and 2016
2008 and 2011

Exercise problem 2:

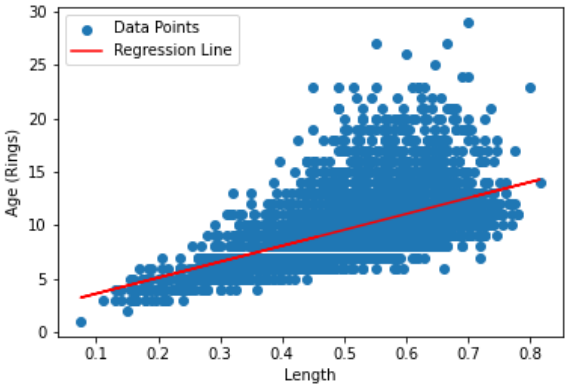


Slope: 0.03475229434762899
Intercept: 51.84358978188418

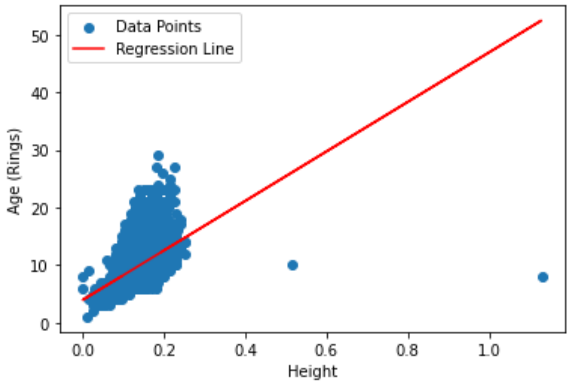
As we can see from the least square regression model, we have a positive correlation between the Employed data and GNP data. The slope of 0.03475~ suggests that each unit increase in GNP correlates to an increase of 0.03475 for the Employed value. The slope value tells us what the value of GNP is, when Employed is 0.

Exercise problem 2:

Length vs Age (Rings)

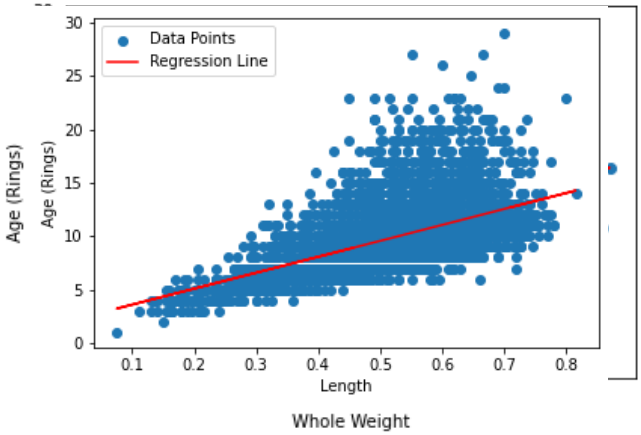


Height vs Age (Rings)

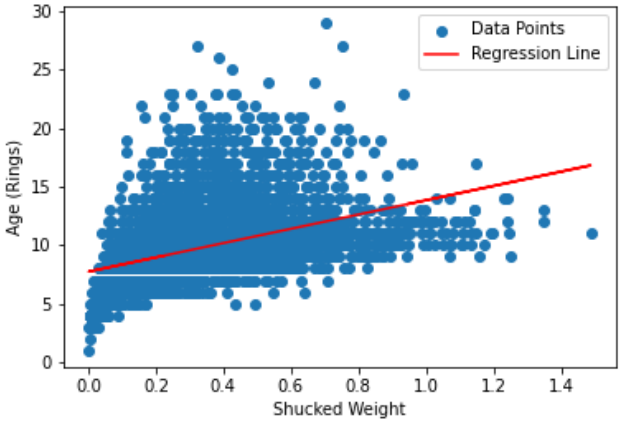


Mean Squared Error (MSE): 7.171674300685373 **Root Mean Squared Error (RMSE): 2.6779981890743265**	**Mean Squared Error (MSE): 7.163015760956357** **Root Mean Squared Error (RMSE): 2.676381094118765**

Whole Weight vs Age (Rings)

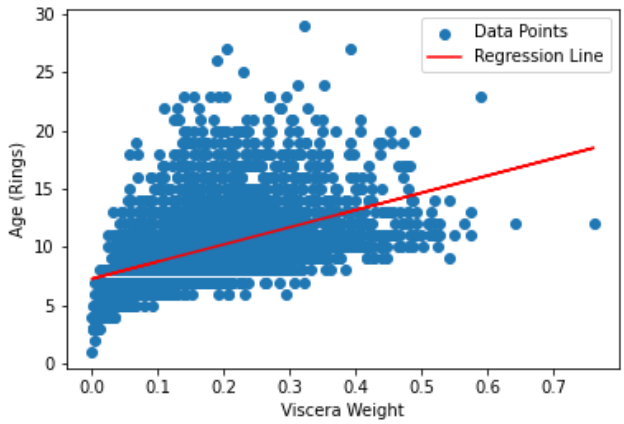


Shucked Weight vs Age (Rings)

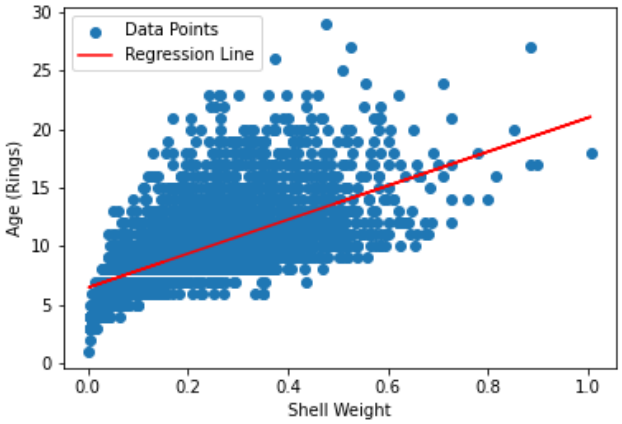


Mean Squared Error (MSE): 7.357868018145953 **Root Mean Squared Error (RMSE): 2.712539035322064**	**Mean Squared Error (MSE): 8.551768957831912** **Root Mean Squared Error (RMSE): 2.924340773205461**

Viscera Weight vs Age (Rings)



Shell Weight vs Age (Rings)



Mean Squared Error (MSE): 7.754738744043551 **Root Mean Squared Error (RMSE): 2.784733154907944**	**Mean Squared Error (MSE): 6.299590441525941** **Root Mean Squared Error (RMSE): 2.5098984922753234**

As we can see from the Multiple linear regression models, there seems to be a positive correlation between each data point and rings. As each statistic increases, the number of rings increases as well. Based on the values of MSE and RMSE, the models seem to be reasonably accurate based on the data provided. One thing that would make the model easier to interpret would be to group the data. This would lead to less data points, but it would be easier to read.