Problem Set 2

Due date: 25 September

Table of contents

Please upload your completed assignment to the ELMs course site (under the assignments menu). Remember to include an annotated script file for all work with R and show your math for all other problems (if applicable, or necessary). Please also upload your completed assignment to the Github repository that you have shared with us. *We should be able to run your script with no errors.*

Total points: 30

Question 1

Points: 5

Using the `states` data, produce a scatterplot of the variables `womleg_2015` and `libpct_m` (with `womleg_2015` as the dependent variable on the y-axis). Describe the scatterplot and include a copy of it. Note any suspected outliers, if any (a visual inspection will suffice for this question). Lastly, give the general equation for the correlation between `womleg_2015` and `libpct_m` (include as much information as possible), but do not solve it.
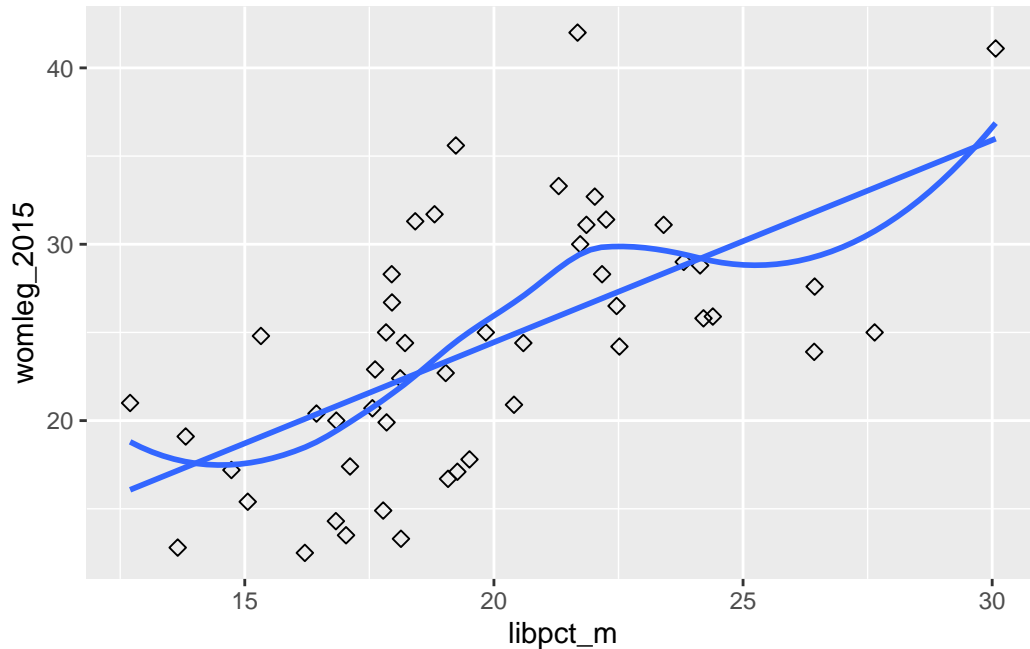
> **i** Note
>
> The `states` data set can be found in `poliscidata::states`. Take a look at `?states` to see what these variables measure.

Ans:

{
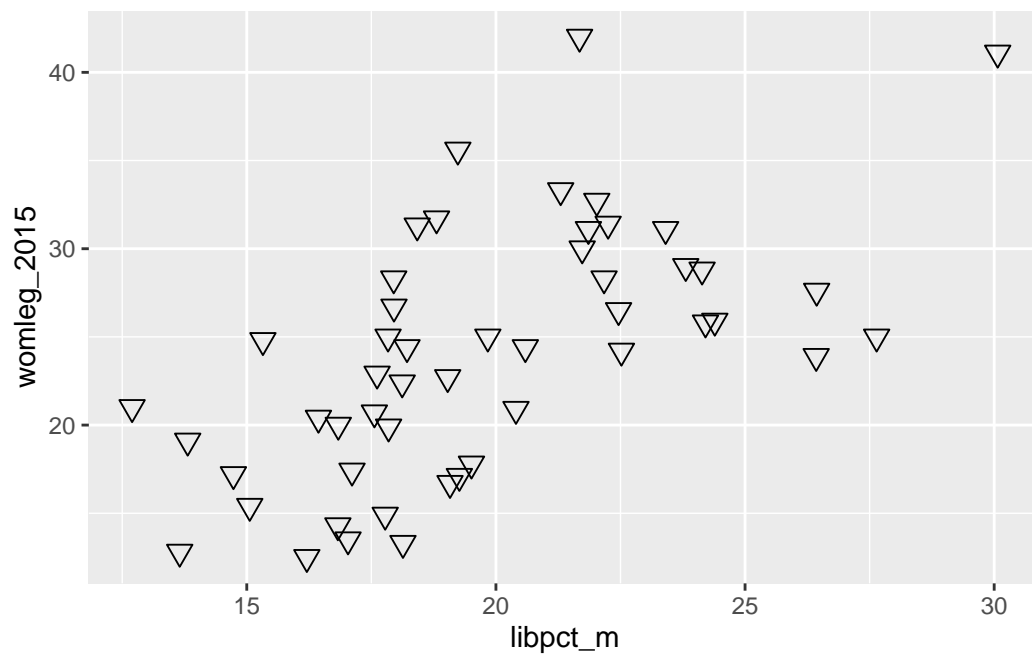
Adding trend lines to determine the strength and the direction of the relationship:

Based on the scatter plot, I have added two trend lines , one linear and the other non-linear to see the variable relationships. Based on the upward movement of lines, we can see that the two variables are positively linked.

Based on a visual of the plot, we can see two points that can be considered outliers that are at the very top of the graph

General equation of correlation:

```
$$
\frac{ \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) }{%
      \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}
$$
```

Pearson correlation is used in order to determine the strength of association between two continuous variables.

```
[1] 0.6088832
```

A correlation of 0.6 is fairly strong. I would assume from the plots and the pearson correlation that these two variables are positively correlated.

	(1)
(Intercept)	1.524
	(4.329)
libpct_m	1.146
	(0.215)
Num.Obs.	50
R2	0.371
R2 Adj.	0.358
AIC	318.4
BIC	324.1
Log.Lik.	−156.205
RMSE	5.50

Question 2

Points: 5

Regress `womleg_2015` (as the dependent variable) on `libpct_m` and report the results in a professionally formatted table. Write the model equation with the estimated coefficients and interpret them. What does the value of R^2 tell us about this model?

Model equation: womleg_2015 = libpct_m * 1.145986 + 1.523990

Interpretation:

The intercept term, in this case 1.523 is the value the model returns when all the variables are zero. This means that if the state has zero liberals, there would be 1.5% of the state legislators being women.

The coefficient for the percentage term is 1.45. This means that on average an increase of percentage of liberals is almost directly related to the percentage of women legislators in 2015.

Although we should note that regression results that are based on percentages should not be directly interpreted in this manner.

R2 value is : 0.371. This means that only about a little more than a third of the variance of the Y variable can be adequately explained by the libpct_m variable.

This would generally not be considered a good model

Question 3

Points: 5

Based on this regression, find the predicted value, the observed value, and compute the residual for the state of Colorado and then the state of Georgia. Lastly, compute the total aggregate error from those two select observations combined (i.e., Colorado and Georgia).

> 💡 **Tip**
>
> Think RSS.

Ans:

```
[1] 42
```

```
        6
26.36757
```

```
        6
15.63243
```

```
[1] 22.9
```

```
       10
21.71097
```

```
       10
1.189031
```

```
[1] 245.7868
```

Question 4

Points: 5

Using the **states** dataset, assess the relationship between the following two variables: **obama_win12** and **gun_rank3**. Construct a cross-tab and describe the nature of the relationship (if any) in detail.

obama_win12		More restr	Mid	Less restr	All
No	N	1	5	18	24
	% row	4.2	20.8	75.0	100.0
Yes	N	14	9	3	26
	% row	53.8	34.6	11.5	100.0
All	N	15	14	21	50
	% row	30.0	28.0	42.0	100.0

> **ℹ Note**
>
> The variable `Obama_win12` is a dichotomous indicator of whether Obama won the state in 2012 (Obama won; Obama lost). The variable `gun_rank3` represents the general (ordinal) extent of gun restrictions in each state (more restrictions; middle restrictions; less restrictions).

> **🔥 Caution**
>
> Please note that you would customarily want a greater number of observations within each cell before conducting such an analysis.

Ans:

We do note that the number of observations are not enough to generate any meaningful understanding. However, with just basing our analysis on the crosstable, we can see that on average Obama won in states that had far greater gun control restrictions. 54% of all states that had high control laws elected Obama. The contrast is even great for less restrictive where 75% of all states with less restrictive laws did not elect him. This suggests are strong relationship between the two variables.

Question 5

Points: 5

I hypothesize that religious identifiers in the mass public are less likely to support federal government support of scientific research. I use data from the General Social Survey to evaluate this hypothesis. In particular, I use a three-category indicator of religious attendance to measure religious identification (low attendance; moderate attendance; high attendance) and a three-category indicator of perceptions toward the federal government's support for scientific research (federal government provides "too little" support; "about right"; federal government provides "too much" support). Complete the cross-tab below so that you may properly evaluate my hypothesis.

Religon		Too Little	About Right	Too Much	All
Low	N	342	356	106	804
	% row	42.5	44.3	13.2	100.0
Moderate	N	190	213	51	454
	% row	41.9	46.9	11.2	100.0
High	N	182	287	91	560
	% row	32.5	51.2	16.2	100.0
All	N	714	856	248	1818
	% row	39.3	47.1	13.6	100.0

> **i** Note
>
> Table entries are raw counts of observations within each cell.

```
Relig.      |        Supporting Scientific Research
Attendance  | Too Little  About Right   Too much |  Total
-----------------------------------------------------------
       Low  |      342         356          106    |

  Moderate  |      190         213           51    |

      High  |      182         287           91    |


-----------------------------------------------------------
     Total  |                                       |
```

Formatted cross table:

Based on the crosstab analysis there does not seem to be a very high correlation between belief in religion and support of science funding. The percentage spread for Low, Moderate and High is roughly the same in relation to science funding. High religion has a slight increase in thinking that there is too much support, but it is comparable to the other categories.

Question 6

Points: 5

dem_adv		More restr	Mid	Less restr	All
high	N	0	4	9	13
	% row	0.0	30.8	69.2	100.0
low	N	6	5	2	13
	% row	46.2	38.5	15.4	100.0
medium	N	11	8	5	24
	% row	45.8	33.3	20.8	100.0
All	N	17	17	16	50
	% row	34.0	34.0	32.0	100.0

Say I wish to explore the relationship between the relative advantage of Democrats (`dem_advantage`) in a state and abortion policy (`abort_rank3`). The `dem_advantage` variable is a continuous indicator where higher values represent a greater Democratic advantage among the mass public; `abort_rank3` is an ordinal indicator for the extent of abortion restrictions in each state (fewer restrictions; middle restrictions; more restrictions). To explore this relationship, complete the following:

Part A

Create a new variable (i.e., `dem_adv`) based on the `dem_advantage` variable. Calculate the summary statistics of `dem_advantage` and assign the following values to our new variable: if `dem_advantage` is less than the first quartile, set `dem_adv` to `Low`; if the value for `dem_advantage` is greater than the first quartile and less than the third quartile, set the value to `Mid`; and if the value of `dem_advantage` is greater than the third quartile, set the value to `High`.

Part B

Create a crosstab using R; include your results in a professionally formatted table.

Part C

What relationship (if any) is there between the relative advantage of Democrats is a given state and the restrictiveness of Abortion policy?

Based on the crosstab data, there seems to be more advantage the less restrictive abortion is in the state. Where the dem advantage is high, zero states have heavily regulated abortion. The opposite is also true strengthening the argument, dems have low and medium advantages in states where the restrictions are high nearly 45% of states in the more restrictive category have less dem advantage.