The Stata Journal

Choosing an appropriate real-life measure of effect size: the case of a continuous predictor and a binary outcome

Ronán M. Conroy
Royal College of Surgeons in Ireland
rconroy@rcsi.ie

Abstract. A case study of data on age and pregnancy is used to point up some morals for practicing data analysts, including the superiority of regression over t tests, exploratory scatterplot smoothing as a key method of checking form of relationship, and the value of logistic regression followed by `adjust` as a way of getting at the numbers of most interest.

Keywords: st0021, adjust, correlation, logistic regression, regression, scatterplot smoothing, t tests.

1   Looking for measures of effect size

I work as a biostatistician in a medical school. Researchers often come looking for p-values. However, there are two reasons why significance tests are less useful than an appropriate measure of effect size with a confidence interval. First, many studies have low statistical power; their biggest risk is not a false-positive conclusion, but a false-negative one. By concentrating on p-values, we draw attention away from study precision. Second, statistical significance does not mean real life importance (and likewise statistical non-significance does not rule out real-life importance). In medicine, it is important to be able to translate data into an informed prediction that we can use to help with decision-making about individual patients. This kind of data analysis is quite distinct from one that tests theoretical hypotheses.

2   The case study: age and pregnancy

In this case, a researcher doing work on the outcome of *in vitro* fertilization has brought some data in for analysis. (*In vitro* fertilization is often known as fertility treatment: eggs are harvested, fertilized in the laboratory, and then re-implanted in the woman.) The researcher is interested in whether the woman's age is correlated with a successful outcome, namely, pregnancy. The least informative thing that I can think of doing here is a correlation, so here is the slightly edited Stata output:

```
. ktau agewoman pregnant
  Number of obs =      341
Kendall's tau-b =      -0.1321

Test of Ho: agewoman and pregnant are independent
     Prob > |z| =       0.0040  (continuity corrected)
```

There is a significant negative correlation between pregnancy and age, but we are virtually no nearer to knowing anything useful about the relationship. Do the chances of pregnancy decline dramatically or slowly with age? We need some measure of the size of the relationship. Because correlations are scaled to have a magnitude between 0 and 1, they cannot easily be converted to real-life units. Like any statistic, Kendall's τ_b can be interpreted as a measure of effect size (in this case, the difference in two probabilities), but its value in explaining things to anyone is pretty close to zero!

3   regress is better than ttest

We could learn more about the ages of the women who did and did not become pregnant by doing a t test.

```
. regress age pregnant
```

agewoman	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
pregnant	-1.36558	.5321081	-2.57	0.011	-2.412229	-.3189305
_cons	34.16245	.2305222	148.20	0.000	33.70902	34.61589

As you can gather from the command I used, a t test is a simple case of regression, where the predictor variable has just two values. The t value in the regression output is the same t that you get from a t test. I could have used Stata's `ttest` command, but `regress` allows me to expand the model to adjust for potential confounding variables. For instance, women with a diagnosis of primary infertility were younger than the other women. I could add this to the model and see if my age difference still holds when adjusted for type of infertility. `regress` encourages you to build models to explain the difference in means, while `ttest` is a dead end, and so I seldom use it. Note also that `regress`, as a built-in Stata command, is faster than `ttest`, implemented as a command defined by an ado-file.

The regression output tells us that the the women who become pregnant were on average 1.36558 years younger than those who did not, with a confidence interval of 2.412229 to 0.3189305. Actually, that number of decimal places is utterly ridiculous in real-life terms. Giving ages to one place of decimals tells us all we need to know: the women who became pregnant were 1.4 years younger (95% CI 2.4 to 0.3). One place of decimals gives us age difference to within about a month; two places of decimals would give it to within a couple of days! This is clearly spurious precision: if we have a margin of uncertainty of 2.7 years around our estimate, why give the estimate itself to two decimal places?

4   Checking the form of the relationship

We have been trying to do something better that we were simply doing wrong. This approach has confused a very basic issue. We are not interested in predicting a woman's age given whether she became pregnant or not; it is the other way around. We want to predict whether she will become pregnant given her age. The response variable in the regression is pregnancy. But what kind of relationship could there be between age and pregnancy? There might be a simple linear one, with chances of pregnancy dropping steadily as women get older. Or, there might be a more curvilinear one, with the relationship steepening at older ages. Or, finally, there might be a threshold effect: chances of pregnancy might be roughly the same until the woman reached a critical age, after which they began to drop significantly. We must have some idea of the shape of the relationship if we are to find an appropriate way of reporting it.

My first step is to draw a scatterplot smoother. I used `autosmoo`, which Peter Sasieni published in STB-41, see Sasieni (1998). It is not a part of "official" Stata, but you will locate it if you use the `findit` command to search for "smoothing". There is also a complementary approach implemented in the `lintrend` command introduced in Garrett (1996).

Figure 1 shows the scatterplot smoother. I have suppressed a small number of patients who are under age 28; the data are too sparse, and make the smoother jump around misleadingly. I have added extra jitter so that the numbers of women in each age group are visible; those who became pregnant are shown above the graph, and those who did not are shown below the graph. I have also made a couple of cosmetic changes to the original output; I wouldn't normally do this, of course, as this is exploratory data analysis.

```
. autosmoo pregnant age if age > 27, log jitter(1) ylab(0) xlab(30,35,40) s(d)
```
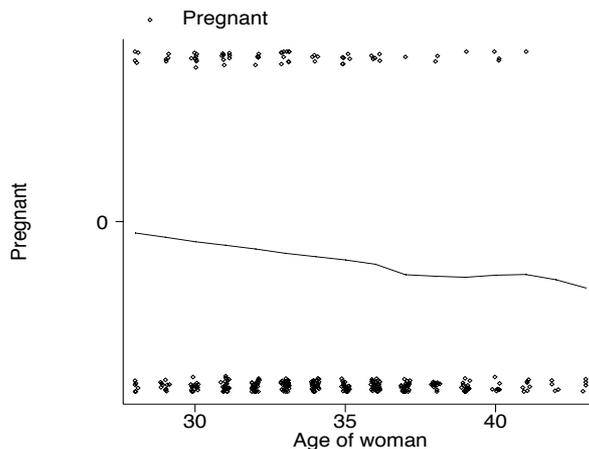
This is reassuring. The chances of pregnancy seem to decline steadily with age. There is no evidence that beyond a certain age they suddenly plummet, and no evidence that the relationship is curved in some way. If the smoother had looked nonlinear, I would have had further work to do, checking to see if I could transform age into a more linear relationship. In the event, a linear model seems to be a useful approximation to the data.

5 From logistic to adjust

The next step, then, is a logistic regression, using age to predict pregnancy:

```
. logistic pregnant age
Logit estimates                              Number of obs   = 341
                                             LR chi2(1)      = 6.55
                                             Prob > chi2     = 0.0105
Log likelihood = -161.37631                  Pseudo R2       = 0.0199
```

pregnant	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
agewoman	.9111197	.0336204	-2.52	0.012	.8475514	.9794557

This is getting closer to an informative model. The odds of pregnancy drop with each passing year by about 9% (1 minus 0.91). However, this tells us how age affects the chances of pregnancy in a relative rather than an absolute way; it does not give us the answer to the question that any couple would ask: given the age of the woman, what are the chances of success?

Here is the result of an `adjust` command, given after the `logistic` command.

```
. adjust if mod(age,5)==0, by(age) pr f(%4.2f) ci
```

```
Dependent variable: pregnant    Command: logistic
```

Age of woman	pr	lb	ub
25	0.34	[0.21	0.49]
30	0.24	[0.18	0.31]
35	0.17	[0.13	0.21]
40	0.11	[0.07	0.18]

```
    Key:  pr        =  Probability
          [lb , ub] =  [95% Confidence Interval]
```

The `adjust` command has given predictions only for those ages that can be divided by five exactly: in other words, the remainder on dividing by 5 is 0, or in Stata terms, `mod(age,5) == 0`.

This brings us to real results in an interpretable form—finally! The predicted success rate is about a third at age 25, is about a quarter at age 30 and falls to about 10% at

age 40. (A success rate of about a third is what fertile couples achieve by regular sexual intercourse, so it's a good rate in absolute terms.) Furthermore, the confidence interval reminds us of the limits of our knowledge. The success rate in 25-year-old women is somewhere between 20% and 50% in round figures, while the rate in 40-year-olds is between 7% and 18%. Confidence intervals guard against a false sense of certainty.

6 Conclusions of the case study

So, the real-life conclusions from the data were:

- Age makes a difference, with chances of success dropping from a third at age 25 to about 10% at age 40.

- There is no evidence of a threshold age beyond which chances of success suddenly fall dramatically (which is good news for couples).

By sharing this information with couples thinking of trying *in vitro* fertilization, the doctor can help them make a more informed decision based on an estimate of their chances of success.

Finally, I should stress that these results apply to one particular clinic. It may be that different clinics have different patient selection criteria, and that even in this clinic they have different selection criteria for older patients. These are issues that I would expect the researcher to address in her paper: they are beyond the scope of statistics, but should not, of course, be outside the awareness of the statistician.

7 Morals for practical data analysis

- `regress` is neater and faster than `ttest` and encourages further exploration of data.

- You can examine the relationship between a continuous predictor and a binary outcome using a scatterplot smoother.

- `adjust` can make predictions that correspond to real-life cases. By showing how your predictor variables change outcome over a typical spectrum of cases, you can help people interpret the importance of your results.

- Correlations are rarely the statistic of choice, except in some cases in which the variables are measured on arbitrary scales, or in which the existence of a relationship is more important than its size. And there are no confounding variables to take into account: that is to say, almost never. In addition, they are a dead end. A regression model can be built up intelligently.

- Each extra decimal digit that you report contains one-tenth the information of the previous one, and makes it harder to see the important figures.

8   References

Garrett, J. M. 1996. sg50: Graphical assessment of linear trend. *Stata Technical Bulletin* 30: 9–15. In *Stata Technical Bulletin Reprints*, vol. 5, 152–160. College Station, TX: Stata Press.

Sasieni, P. 1998. gr27: An adaptive variable span running line smoother. *Stata Technical Bulletin* 41: 4–7. In *Stata Technical Bulletin Reprints*, vol. 7, 63–68. College Station, TX: Stata Press.

About the Author

Ronán Conroy is a biostatistician at the Royal College of Surgeons in Ireland.