**GEOG/ES&P 330: Quick Guide to Effective Statistical Reporting (focussing on Chi-square)**

**Remember: significance is not the same as strength, and neither one is the same as power!!!**

Can you safely reject the NULL hypothesis? That is, **are your results significant?** This means what is the probability of making a Type 1 error (a false positive finding) if you reject the null hypothesis?

> Before doing anything, settle on an ***alpha* standard** to define "safely reject the null hypothesis" and think through why you picked a particular standard. This entails imagining the worst thing that could happen if you make a Type 1 error and what the worst thing that could happen if you make a Type 2 error. You want to limit the scarier error (and explain your reasoning).
>
> > If your main concern is avoiding a false positive (Type 1 error), *alpha* needs to be small (0.05 or even 0.01 if a false positive could be dangerous). Getting excited over nothing is usually the scarier thought for a scientist, so the **0.05 standard is the most common** balance of terrors.
> >
> > If you are more concerned about avoiding a false negative (Type 2 error), *alpha* can be more generous (*e.g.,* 0.10). This may be appropriate for pilot studies or small. exploratory studies, if you're afraid of missing something that is there.
>
> **Calculate the p-value** (or prob-value or significance). The prob-value has to be below your *alpha* standard for your results to be significant: $\mathbf{p < \alpha}$. With p, small is beautiful.
>
> Alternatively (as part of calculating prob-value), calculate the observed test statistic (such as $\chi^2_{calc}$) and compare it with the critical test statistic (such as $\chi^2_{crit}$) for your given *alpha* standard. Your results are significant if the calculated test statistic is bigger than the critical test statistic: $\boldsymbol{\chi^2_{calc} > \chi^2_{crit}}$. With calculated test statistics, the more the merrier.

**How strong is the effect you've discovered?** How much do the data differ from the null hypothesis? How much of the variation in one variable is explained by variation in the other variable? How dramatic is the effect? That is given by effect-size measures, such as $R^2_{adj}$ in regression analysis, Cohen's D in t-tests, and Cramèr's V in Chi-square. These are generally set to vary from 0.0 to 1.0: The closer to 0, the weaker the effect; the closer to 1, the stronger the effect. Interpretation of these scores varies among authors and how big the table is (see https://www.statology.org/interpret-cramers-v/ ). To be conservative, **let's use <0.10 for faint or trivial; 0.10-0.29 for weak; 0.30-0.49 for moderate; and ≥0.50 for strong**.

Last, figure out **how much statistical power your study achieved**. Power is the ability to minimize making a Type 2 error (of dismissing a small effect as not significant when, in fact, it is real). It is possible to have no significant results and a large effect size, which is really baffling. This can happen if your study sample size is just too small to detect that effect size as significant. It's always ideal to try to figure out how large a sample to collect ahead of time to be able to detect a given effect size at a particular *alpha* standard but sometimes we don't get that luxury and you have to run with what you've got. In that situation (which describes your projects), all you can do is report the power your sample achieved after the fact. **Ideally, power should be ≥0.80**. So, if you have no significant results AND the effect size is ≥0.20, look at power. If it's <0.80, it's possible you just don't have enough data and you should recommend more be collected to see if a larger sample can push a given effect into significance.

Remember: You can get very significant results with lots of power and yet have a tiny effect size (small, but real). You can also get a large effect size but no significance if you don't have enough data (the effect may be real, who knows? but your study has too little power to show it).