Does Humidity Cause Death by Heatstroke? Important Evidence from the Cincinnati Sunstroke Epidemic of 1881

Are you worried that the sun will kill you off?

Are you a serious researcher of death by sunstroke?

Are you obsessed with 1881?

No? Well, you've come to the right place. Here we're going to poke fun at the work some researchers did way back in the day, using modern statistics, to see if their claims are actually true.

Motivation

I stumbled across this mildly interesting paper (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2272406/) about the Cincinnati Sunstroke Epidemic of 1881, and two things caught my attention. (Miles 1881) First, there were two pages tabulating all of the data. Second, the author made an interesting claim that I had to follow up on. He claims:

"The object in thus presenting these tables is to prove by the observations, that the *dryer the atmosphere* the greater will be the increase in the death-rate from insolation."

Well, I've seen those tables and they're pretty dang convincing.

Humidity (saturation being 100).	Number of Days.	Number of Deaths.	Rate per Day.
Between 40 and 50	3	89	29⅔
Between 50 and 60	18	174	9.66⅔
Between 60 and 70	8	21	2.625
Between 70 and 80	3	3	1.
Between 80 and 90	1	1	1.

It appears that as humidity decreases, there rate of people dying off, per day, grows exponentially.

So, convincing tables aside, how do these results hold up to, say Ordinary Least Squares? (If you don't know what that is, it's a simple line of best fit. Intuitively, we're drawing the line that fits the data best). If he's right, then we would expect there to be a large negative relationship between the two. I ran a regression, and the results are not so good.

```
Linear regression                               Number of obs     =         81
                                                F(1, 79)          =       2.40
                                                Prob > F          =     0.1252
                                                R-squared         =     0.0319
                                                Root MSE          =     9.5932
```

deaths	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
humidity	-.1930884	.1246089	-1.55	0.125	-.4411161	.0549394
_cons	14.72155	8.100128	1.82	0.073	-1.401356	30.84445

Yes, the coefficient is negative, but the probability that there is no effect and he got this by chance is 12.5%

OK, but maybe I'm looking at the problem wrong. He claims that "the thermometer being the same, a dry atmosphere is much more

conducive to sunstroke than a hot and moist atmosphere", so perhaps if we control for temperature, we'll be able to see the actual results.

Linear regression

Number of obs	= 81
F(4, 76)	= 4.86
Prob > F	= 0.0015
R-squared	= 0.3688
Root MSE	= 7.8975

deaths	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
max_temp	-.0001567	.0001481	-1.06	0.293	-.0004518	.0001383
min_temp	.943927	.4226683	2.23	0.028	.1021102	1.785744
mean_temp	.0482279	.2833275	0.17	0.865	-.5160678	.6125236
humidity	-.2481015	.1348488	-1.84	0.070	-.5166761	.0204732
_cons	-52.86712	13.52641	-3.91	0.000	-79.8073	-25.92694

OK, so that's a little better. Maybe there's something to this. Our P statistic (0.07) is still above the traditional benchmark 0.05, but it is close.

What about autocorrelation? There's several ways to check this, but the easiest is just to bootstrap, and see what happens.

Wald chi2(4)	= 16.21
Prob > chi2	= 0.0028
R-squared	= 0.3688
Adj R-squared	= 0.3356
Root MSE	= 7.8975

deaths	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
max_temp	-.0001567	.2352006	-0.00	0.999	-.4611414	.460828
min_temp	.943927	.4707313	2.01	0.045	.0213107	1.866543
mean_temp	.0482279	.3419248	0.14	0.888	-.6219325	.7183883
humidity	-.2481015	.1419468	-1.75	0.080	-.526312	.0301091
_cons	-52.86712	13.40632	-3.94	0.000	-79.14303	-26.59121

Well, it looks like we were only underestimating the standard error by 0.08-ish. That's good.

What about other variables? If the effect we're observing for humidity is actually a causal effect, it shouldn't change very much when we control for more stuff. We have a bunch more weather variables, so let's take a look:

Linear regression

Number of obs	=	79
Replications	=	50
Wald chi2(8)	=	15.76
Prob > chi2	=	0.0459
R-squared	=	0.4008
Adj R-squared	=	0.3323
Root MSE	=	8.0098

deaths	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
max_temp	-.0002943	.184318	-0.00	0.999	-.361551	.3609625
min_temp	1.086697	.4635803	2.34	0.019	.1780964	1.995298
mean_temp	-.0434756	.376717	-0.12	0.908	-.7818273	.6948761
barometer	-2.333888	8.044249	-0.29	0.772	-18.10033	13.43255
humidity	-.3784681	.1817622	-2.08	0.037	-.7347154	-.0222208
rainfall	6.404481	2.898475	2.21	0.027	.7235754	12.08539
cloudy	2.164618	2.756357	0.79	0.432	-3.237743	7.566978
clear	1.991897	2.436563	0.82	0.414	-2.783679	6.767473
_cons	20.04115	244.2314	0.08	0.935	-458.6436	498.7259

Wow, look at that! When we controlled for everything else, humidity became statistically significant.

OK, but there's something very very wrong about this regression. Look closely. Found it? No, it's not that max temperature is statistically insignificant (it would be if we removed min temperature). Apparently, rain kills people. By sunstroke.

If you scroll through the data, this relationship makes sense (not intuitively, but with the data). They happened to get a rainstorm right after the big heat wave that killed a bunch of people off.

OK OK. Lower your pitchforks. I don't think that rain itself actually kills people, especially by sunstroke. There's a set of variables that's killing people that we haven't included in our regression yet—how the weather was yesterday. Let's take a look:

humidity						
--.	-.2282117	.1469565	-1.55	0.126	-.5227191	.0662957
L1.	-.2775118	.1705435	-1.63	0.109	-.6192886	.0642649
rainfall						
--.	4.227987	3.647649	1.16	0.251	-3.082064	11.53804
L1.	6.507176	2.586882	2.52	0.015	1.322948	11.6914

Sorry, this isn't the whole printout for the regression, but you can see the important results. Once we control for the weather the day before, rainfall today isn't killing anyone. And once we control for the day before,

humidity						
--.	-.1405227	.1361199	-1.03	0.307	-.4139276	.1328821
L1.	-.1614105	.1456226	-1.11	0.273	-.4539021	.1310811
L2.	-.2485908	.1579516	-1.57	0.122	-.565846	.0686644
rainfall						
--.	1.998046	3.466176	0.58	0.567	-4.963973	8.960065
L1.	1.512715	2.758838	0.55	0.586	-4.028575	7.054005
L2.	1.655875	8.057352	0.21	0.838	-14.52779	17.83954

rainfall isn't killing anyone at all. As you can see, the effect of humidity phases out as well (in statistical significance at least). In fact the only significant variable from this last regression is temperature from 2 days ago. (You'll just have to trust me on that one, or download the data and run it for yourself)

Since our coefficient for humidity evaporated once we controlled for the day before, it's probable that it doesn't have a causal impact on sunstroke deaths. It's much more likely that people were dying from sunstroke from prolonged exposure to high temperatures, and the 'effects' we were observing with humidity were because those days happened to be dry ones.

One final note from Google:

If you Google "Does Heatstroke Affect Humidity" you'll get a lot of results all with titles like "Humidity Matters" and "It Really is the Humidity". If you bother to click any, you'll discover the overwhelming consensus: Humidity helps cause heatstroke, but in the opposite direction of what this paper was claiming. This paper claimed that less humidity causes more heatstrokes, but the truth is that humidity causes MORE heatstrokes. If we believed the 'convincing' tables above, we would be factually wrong.