WOW, DEEP LEARNING IS
SOOO...
FUCKING USELESS
memegenerator.net

Figure:

Statistics with Julia

“.....Can he do it on a cold, wet Wednesday night in Stoke?”

Statistics with Julia

1st Year Exploratory Data Analysis, Summary Statistics, Probability, Graphical Methods

2nd Year Hypothesis Testing, Confidence Intervals, Probability Distributions, Linear Models

3rd Year ANOVA and Experimental Design, Residuals, Chi Squared, Stepwise Regression

4th Year PCA, Clustering, Logistic Regression

Statistics with Julia

What is like to teach statistics vs **What it should be like**

The Future according to Kevin

- Remove Pen and Paper Calculations
  (Keep a few good ones)
- Sort out Bad Instructional Design
  (time is short...dont waste time with stupid crap)
- Why the flip is this still an exam question? Is this still the 30s?
- (The t-test is actually fairly robust to non-normality).
- e.g. Sum of Squares Identities in Experimental Design..**BY HAND...F.R.O!!!**

Statistics with Julia

Writing Stats Exams!!

- Copy and Paste questions some past papers
- change a few numbers here and there
- Transforms weights of dogs into heights of cats

$$X \sim 1000, 25^2$$

- Why fix that equation? too much like work?

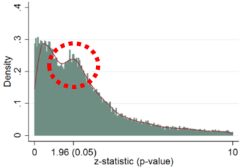
Exam papers take time.... Hey, you got better things to do!!!

- Put in more p-values...but learning to critique the analysss proplerly
- Tell them about P-hacking
- and anyway...What exactly is a confidence interval (for mean?)

Economics

Brodeur et al (AEJ:A, in press)

“Star Wars: The empirics strike back”

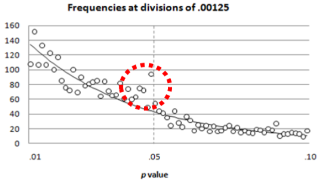


(b) De-rounded distribution of z-statistics.

Psychology

Masicampo Lalande (QJEP, 2012)

“A peculiar prevalence of p values just below .05”



Biology

Head et al (PLOS Biology 2015)

“Extent and Consequences of P-Hacking in Science

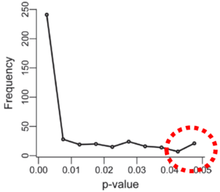
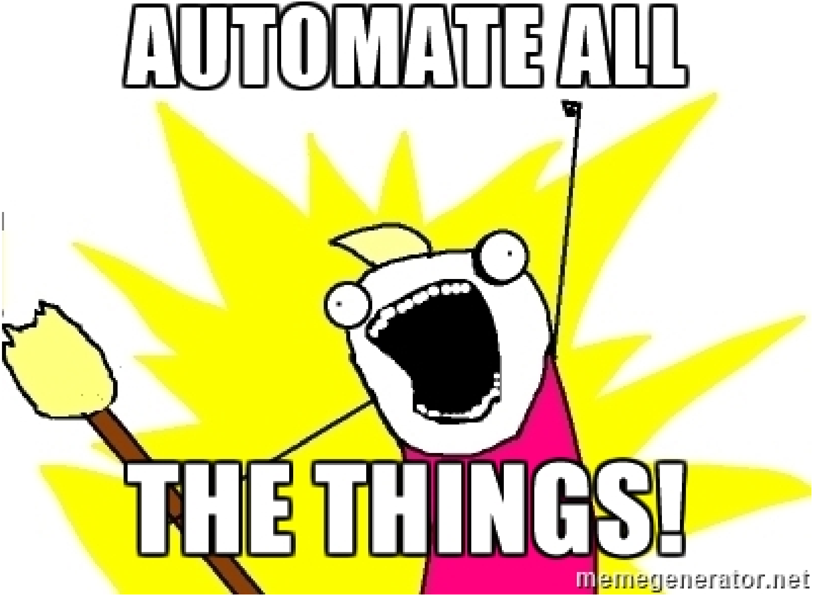


Figure:

Statistics with Julia

- Never omit "Type I" Error and "Type II" Error
- HT is not about what is true or false it is above what you can prove (back up with a sufficient amount of evidence)
- You'd be surprised about how many people dont know that.


AUTOMATE ALL
THE THINGS!
memegenerator.net

Open source, interactive data science and scientific computing

Figure:

Statistics with Julia

Surely students could handle some code?

- `sample()`
- `mean()`
- `t.test()`

They dont have to like it, but they would prefer having to do some basic computing as opposed to.....

Statistics with Julia

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij}^{\,2} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(Y_{ij}-\widehat{Y}_i\right)^2$$

$$= \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(Y_{ij}-\overline{Y}_{i\bullet}\right)^2}_{\text{(sum of squares due to pure error)}} + \underbrace{\sum_{i=1}^{n} n_i\left(\overline{Y}_{i\bullet}-\widehat{Y}_i\right)^2.}_{\text{(sum of squares due to lack of fit)}}$$

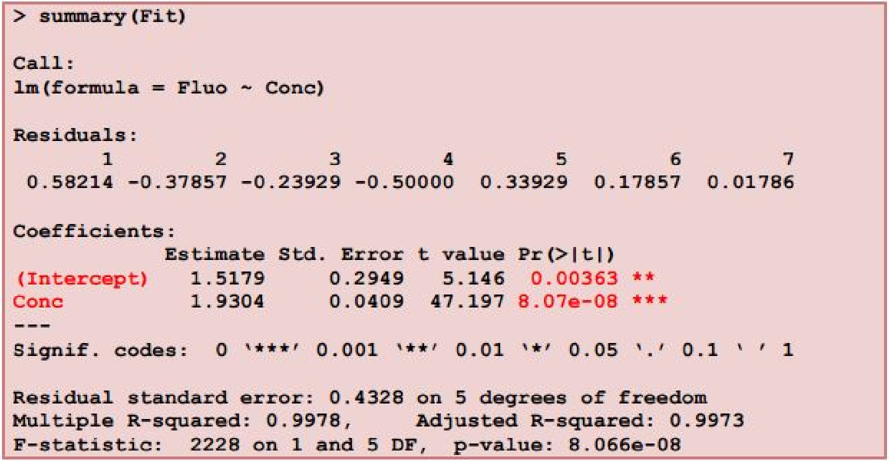
2 Hours of my life wasted!!!
(Although some are worth keeping)

Statistics with Julia

Non-parametric statistics

- ranked data
- Likert scale
- **carrying out a heart transplant with a shovel**
- hard to prove anything to satisfactory degree

Statistics with Julia



```
> summary(Fit)

Call:
lm(formula = Fluo ~ Conc)

Residuals:
       1        2        3        4        5        6        7 
 0.58214 -0.37857 -0.23929 -0.50000  0.33929  0.17857  0.01786 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.5179     0.2949   5.146  0.00363 ** 
Conc          1.9304     0.0409  47.197 8.07e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.4328 on 5 degrees of freedom
Multiple R-squared: 0.9978,     Adjusted R-squared: 0.9973 
F-statistic:  2228 on 1 and 5 DF,  p-value: 8.066e-08
```

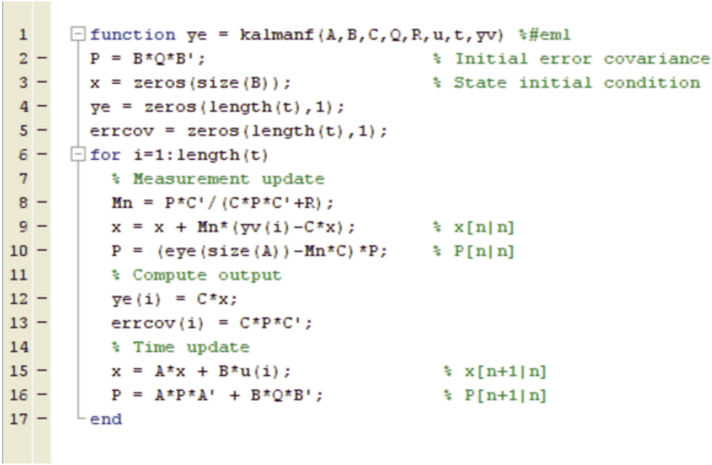
Figure:

Statistics with Julia

This is `R`, but same argument applies to Julia.

- `AIC()`
- `summary()`
- `cor.test()`
- `plot(Fit)`

```
julia> # Pkg.clone("git://github.com/JuliaQuant/MarketData.

julia> using MarketData, Gadfly, HypothesisTests

julia> dist = percentchange(cl).values;

julia> funkydist = dist[100:300];

julia> SignTest(funkydist)
Sign test

median = 0.0
x = 111
n = 201

Two-sided p-value:
p = 0.15816534520094128
```



```
function ye = kalmanf(A,B,C,Q,R,u,t,yv) %#eml
P = B*Q*B';                          % Initial error covariance
x = zeros(size(B));                  % State initial condition
ye = zeros(length(t),1);
errcov = zeros(length(t),1);
for i=1:length(t)
  % Measurement update
  Mn = P*C'/(C*P*C'+R);
  x = x + Mn*(yv(i)-C*x);            % x[n|n]
  P = (eye(size(A))-Mn*C)*P;         % P[n|n]
  % Compute output
  ye(i) = C*x;
  errcov(i) = C*P*C';
  % Time update
  x = A*x + B*u(i);                   % x[n+1|n]
  P = A*P*A' + B*Q*B';                % P[n+1|n]
end
```

Figure:

Statistics with Julia

- StatsBase
- DataFrames
- RDatasets

- Stuff that gets me
- I still think in "R", not MATLAB
- Why is this not working (in Julia)?

```
myData[myData < 400]
```

If I thought in MATLAB, Julia is fairly easy to pick up

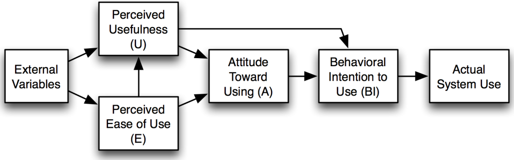

External Variables
Perceived Usefulness (U)
Perceived Ease of Use (E)
Attitude Toward Using (A)
Behavioral Intention to Use (BI)
Actual System Use

Figure: