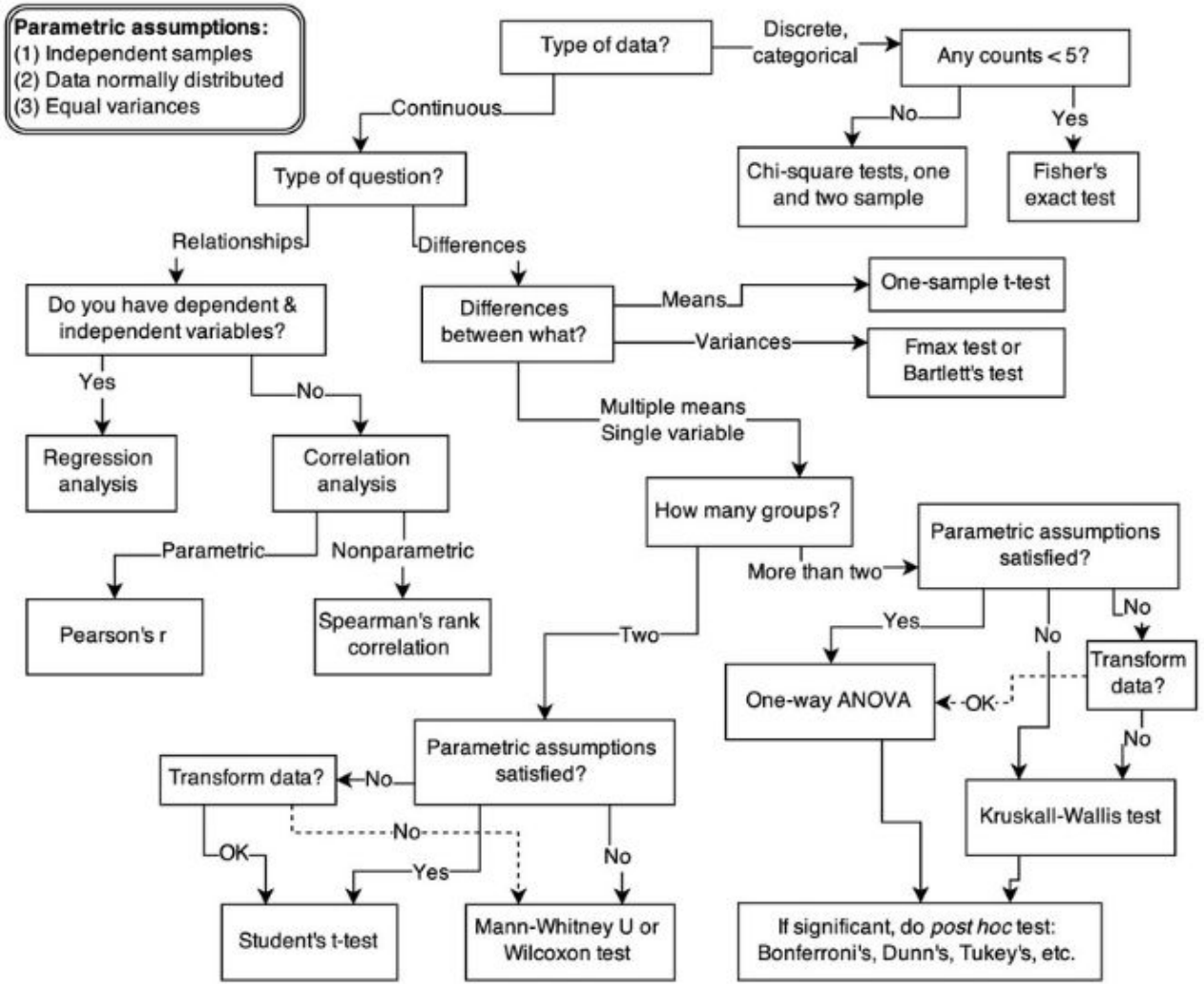
Bayesian 1&2 McElreath

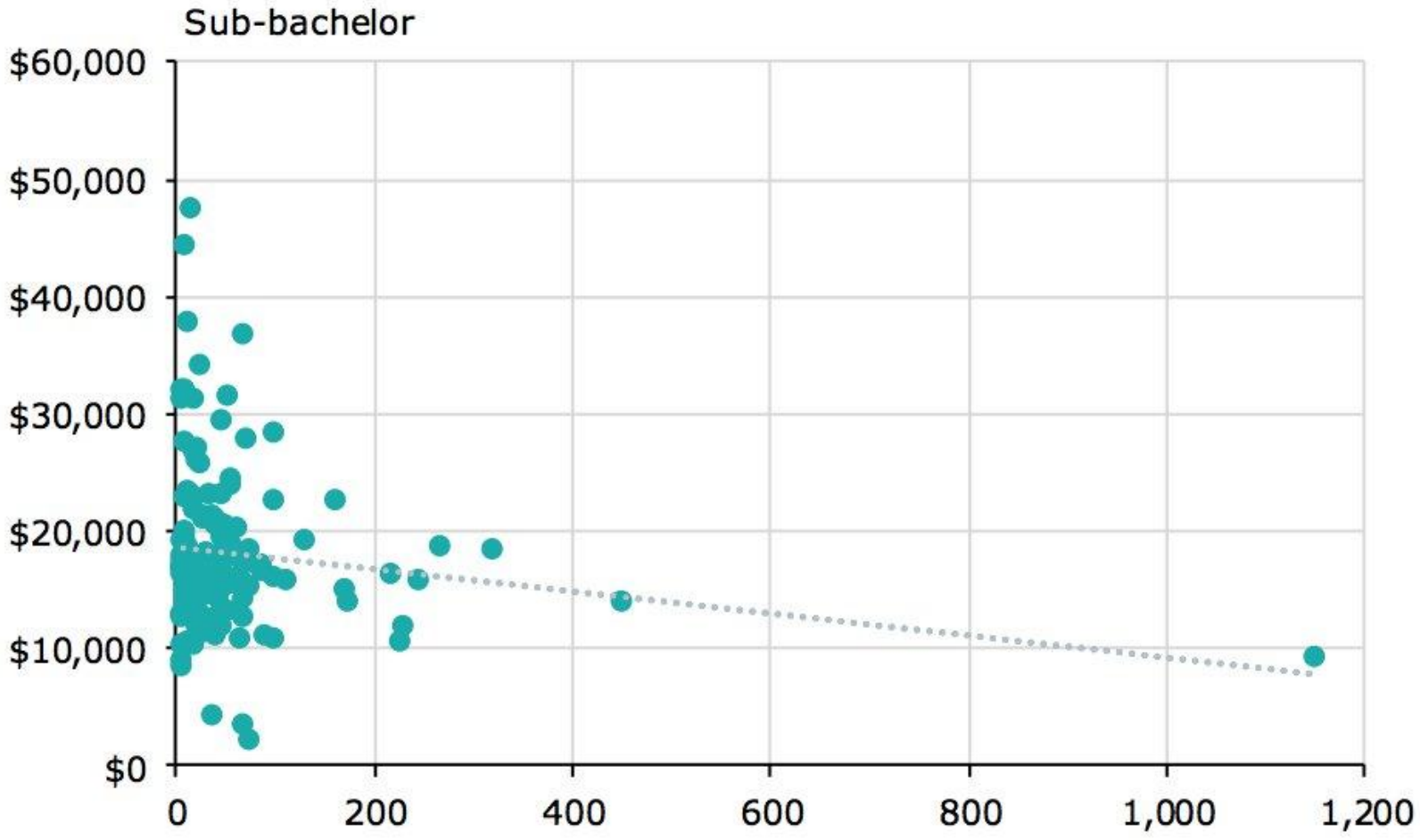
Bayesian 1

Background

Background

1. Science problems are irreducible to statistical test.
2. Popular use of Statistics in 18th centuries: agriculture, R.A Fisher.
    a. Hard problem, but simpler compare to problems in other science
    b. Randomized Control Experiment in Agriculture
    c. Use pesticide or not?
    d. Anova T-test, is the mean differ from zero?


Parametric assumptions:
(1) Independent samples
(2) Data normally distributed
(3) Equal variances
Type of data?
Discrete, categorical
Any counts < 5?
No
Yes
Chi-square tests, one and two sample
Fisher's exact test
Continuous
Type of question?
Relationships
Differences
Do you have dependent & independent variables?
Yes
No
Regression analysis
Correlation analysis
Parametric
Nonparametric
Pearson's r
Spearman's rank correlation
Differences between what?
Means
One-sample t-test
Variances
Fmax test or Bartlett's test
Multiple means
Single variable
How many groups?
Two
More than two
Parametric assumptions satisfied?
Yes
No
One-way ANOVA
OK
Transform data?
No
Kruskall-Wallis test
Parametric assumptions satisfied?
Transform data?
No
OK
No
Yes
No
Student's t-test
Mann-Whitney U or Wilcoxon test
If significant, do post hoc test: Bonferroni's, Dunn's, Tukey's, etc.


Sub-bachelor
$60,000
$50,000
$40,000
$30,000
$20,000
$10,000
$0
0
200
400
600
800
1,000
1,200

Background

Statistical tests are not the answer

The Golem of Prague

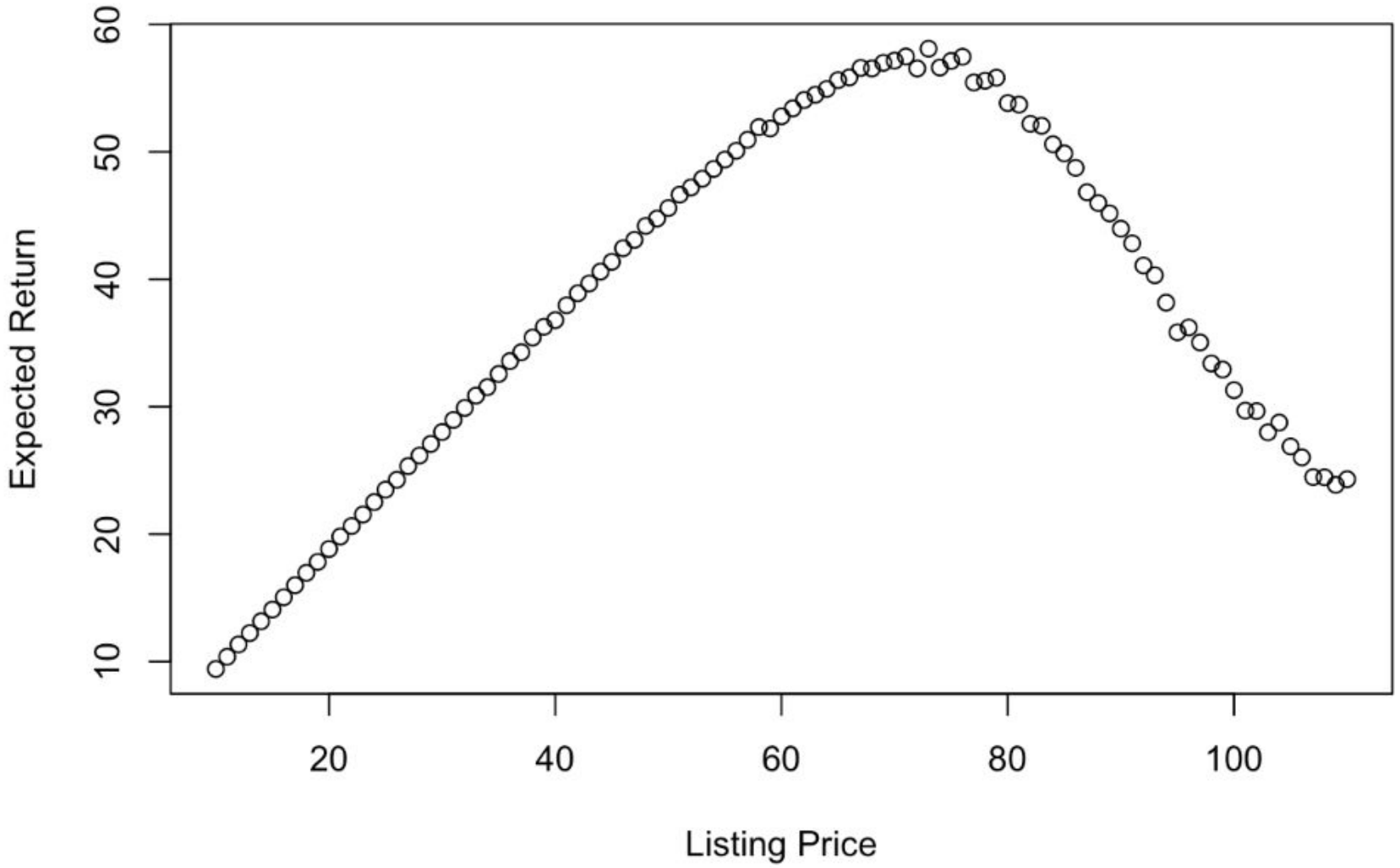
"Even the most perfect of Golem, risen to life to protect us, can easily change into a destructive force. Therefore let us treat carefully that which is strong, just as we bow kindly and patiently to that which is weak."

Rabbi Judah Loew ben Bezalel (1512–1609)

From *Breath of Bones: A Tale of the Golem*

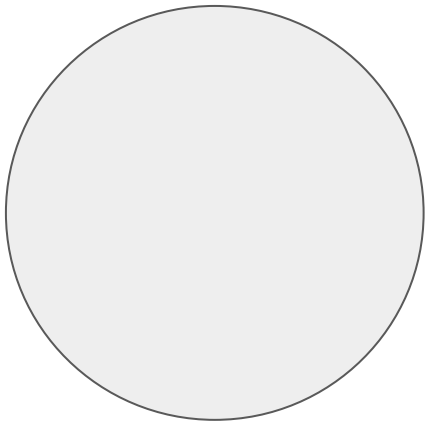
Background

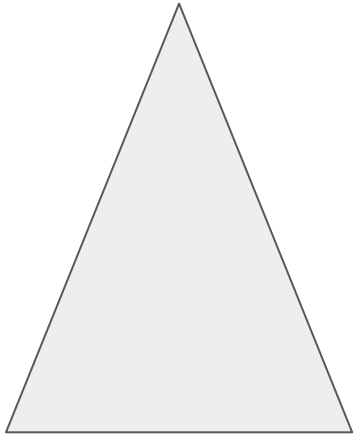
1. Healthy dose of data/statistical model skepticism is good.
2. Enough confidence to be comfortable with confusion
3. Inference should be about decision:
    a. http://www.statsathome.com/2017/10/12/bayesian-decision-theory-made-ridiculously-simple/

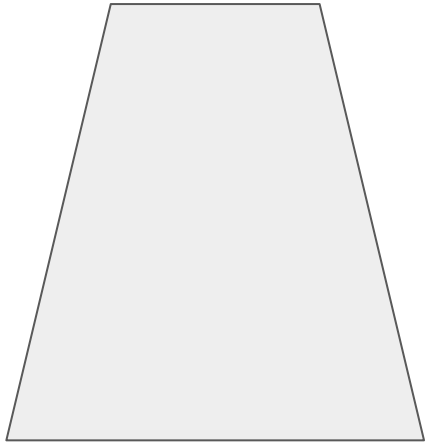

Expected Return
10
20
30
40
50
60
20
40
60
80
100
Listing Price

Background

1. Statistical test is not hypothesis testing







Background

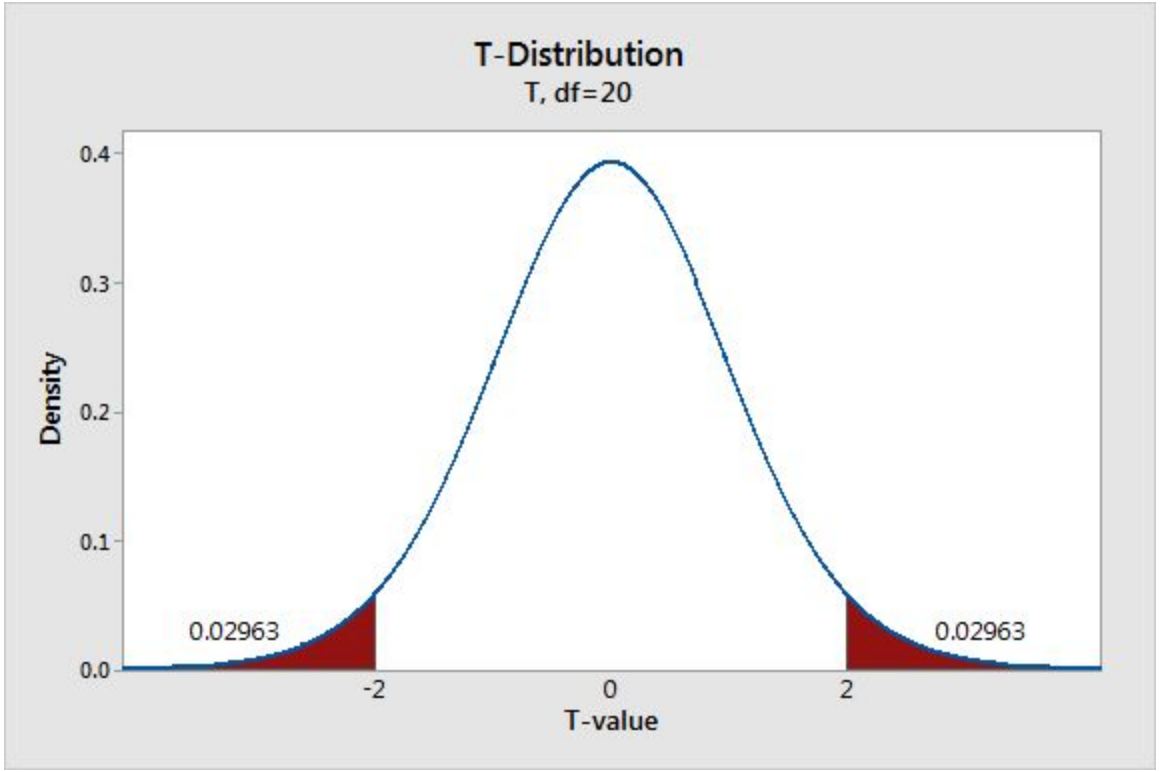
Falsification

“Should falsify the explanatory variable, not the null hypothesis”

https://philosophy.stackexchange.com/questions/23087/what-differentiates-the-scientific-method-from-other-methods

H0: Tidak ada efek obat nyamuk terhadap kesehatan nyamuk

H1: Ada efek obat nyamuk ke kesehatan nyamuk


T-Distribution
T, df=20
Density
0.4
0.3
0.2
0.1
0.0
0.02963
0.02963
-2
0
2
T-value

Golem Engineering

Bayesian Data Analysis

- Use probability to describe uncertainty
- Extends ordinary logic (true/false = testing hypothesis) to continuous plausibility = posterior dist
  - Computationally hard
  - Markov chain Monte Carlo (MCMC) to the rescue
- Used to be controversial
  - Ronald Fisher: Bayesian analysis “must be wholly rejected.”

Multilevel models

- Models with *multiple levels* of uncertainty
  - Replace parameters with models
- Common uses
  - Repeat & imbalanced sampling
  - Study variation
  - Avoid averaging
  - Phylogenetics, factor and path analysis, networks, spatial models
- Natural Bayesian strategy

Multilevel Modeling

- Provinsi
- Sekolah
- Siwa

Cat A:

AA

AB
AC

Cat B

Model Comparisson

- Build multiple meaningful model
- Compare their best fit and level of your assumption acceptance

Occam’s simplification is dangerous, don’t bet your understanding of simplification prior.

The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time

http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf