Statistical Hypothesis Testing

Statistical Hypothesis Testing

- definitions
  - what is it?
- theories
  - Why are we using it?
- Examples
  - How are we using it?
- Applications
  - How is it related to us?

What is it?

- A statistical test provides a mechanism for making quantitative decisions about a process or processes. The intent is to determine whether there is enough evidence to "reject" a conjecture or hypothesis about the process.
  - NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Statistical Hypothesis Testing

- Statistical
- Hypothesis
- Testing

Statistical Hypothesis Testing

- statistical model & statistical inference
  - random variables X
  - distribution P_θ (at least partly unknown)
  - a set of observations/measurement

Statistical Hypothesis Testing

- statistical model & statistical inference
  - "Statistical inference is concerned with methods of using this observational material to obtain information concerning the distribution of X or the parameter ϴ with which it is labeled"
  - "A statistical inference is a procedure that produces a probabilistic statement about … a statistic model"

Lehmann, E.L. Testing Statistical Hypotheses
DeGroot M.H., Schervish M.J. Probability and Statistics

Statistical Hypothesis Testing

- Sample vs. Population
- Example ("Lake Wobegon")
  - Someone claims kids at Lake Wobegon have above average intelligence
  - random samples of 9 kids there with test result of {116, 128, 125, 119, 89, 99, 105, 116, 118}
  - Wechsler scores (the test they take) are scaled to be normally distributed with a mean of 100 and standard deviation of 15.
  - mean of sample: 112.8

http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Statistical Hypothesis Testing

- Null Hypothesis vs. Alternative Hypothesis
- Null hypothesis being “attacked”
- Usually more emphasis on Alternative Hypothesis
  - “Formulate the null hypothesis H_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H_α (commonly, that the observations show a real effect combined with a component of chance variation).”

Weisstein, Eric W. "Hypothesis Testing." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/HypothesisTesting.html

Statistical Hypothesis Testing

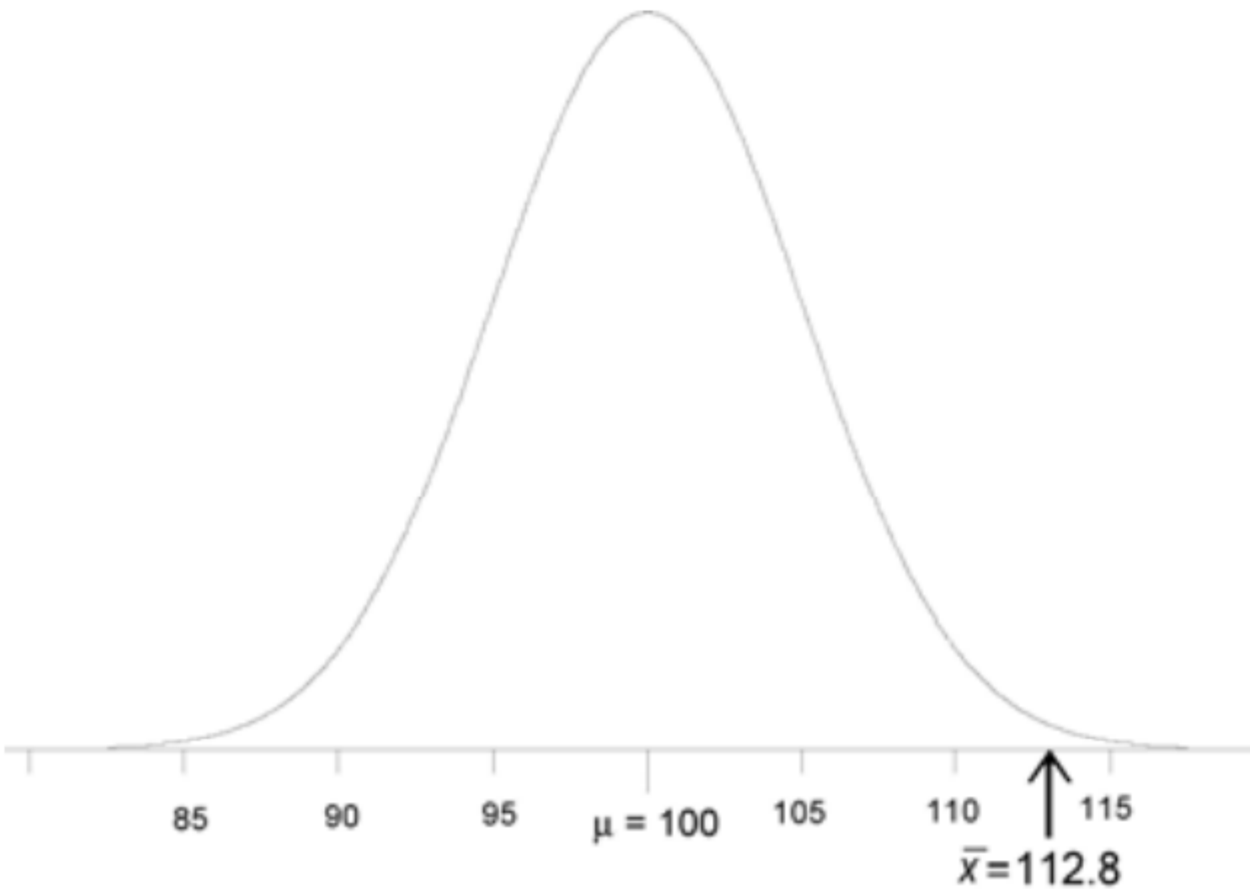
- The idea is to refute H_0 if the sample is statistically far from the population/distribution, which is modeled only under the assumption that H_0 is true.
- we want to know the confidence level of refuting H_0
- Failed to reject the null hypothesis does not mean the null hypothesis is true

Statistical Hypothesis Testing

- one-sided vs. two-sided alternatives
- one-sided:
  - H_0: $u \leq 100$
  - H_a: $u > 100$
- two-sided:
  - H_0: $u = 100$
  - H_a: $u \neq 100$

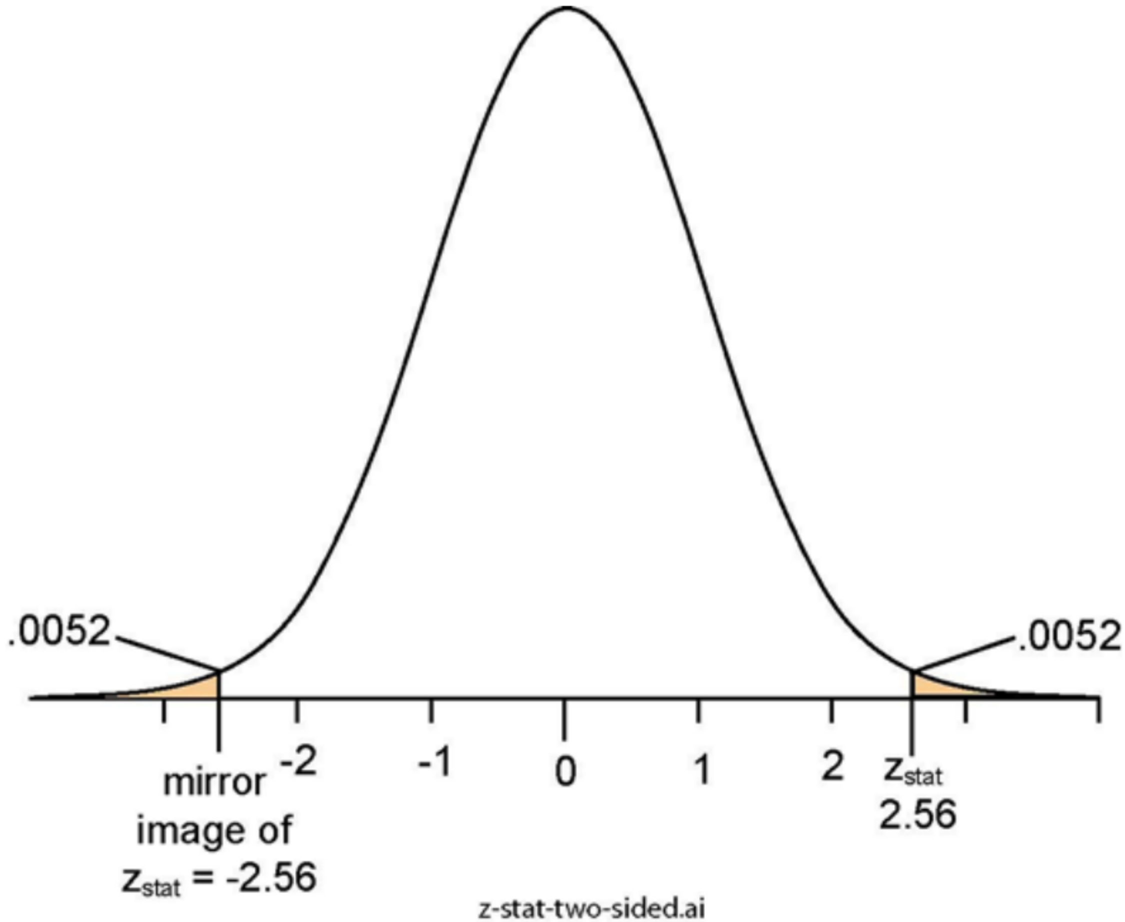
http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Statistical Hypothesis Testing



http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Statistical Hypothesis Testing



http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Statistical Hypothesis Testing

- Test Statistic:
  - The test statistic is a statistical method based on the specific hypothesis test.
  - $T = r(X)$ where X is the random sample from the distribution.
  - H_0 will be rejected if $T \in R$
- Critical Region:
  - the set $S_1 = \{x : r(x) \in R\}$

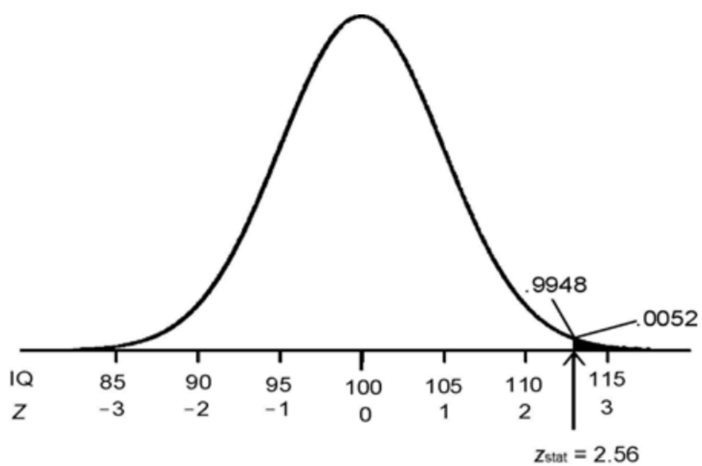
Statistical Hypothesis Testing

- Significance Level: α
  - "A value of $\alpha = 0.05$ means that we inadvertently reject the null hypothesis 5% of the time when it is in fact true. "
- Power: $1 - \beta$
  - "the probability of accepting the null hypothesis when the alternative hypothesis is, in fact, true, is called β."
- often referred as Type 1 and Type 2 errors, respectively

NIST/SEMATECH e-Handbook of Statistical Methods

Statistical Hypothesis Testing

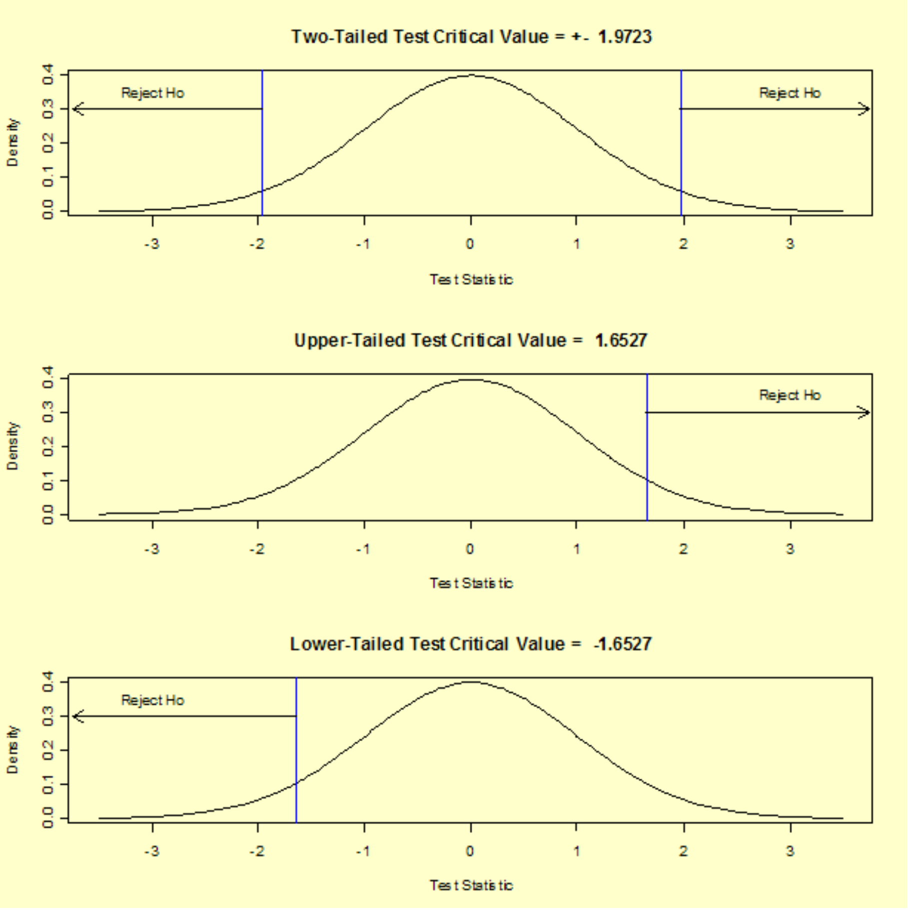
- Critical Region: encompasses those values of the test statistic that lead to a rejection of the null hypothesis.
- P-value: the probability that a test statistic at least as significant as the one observed



z-stat-one-sided.ai

NIST/SEMATECH e-Handbook of Statistical Methods

Statistical Hypothesis Testing



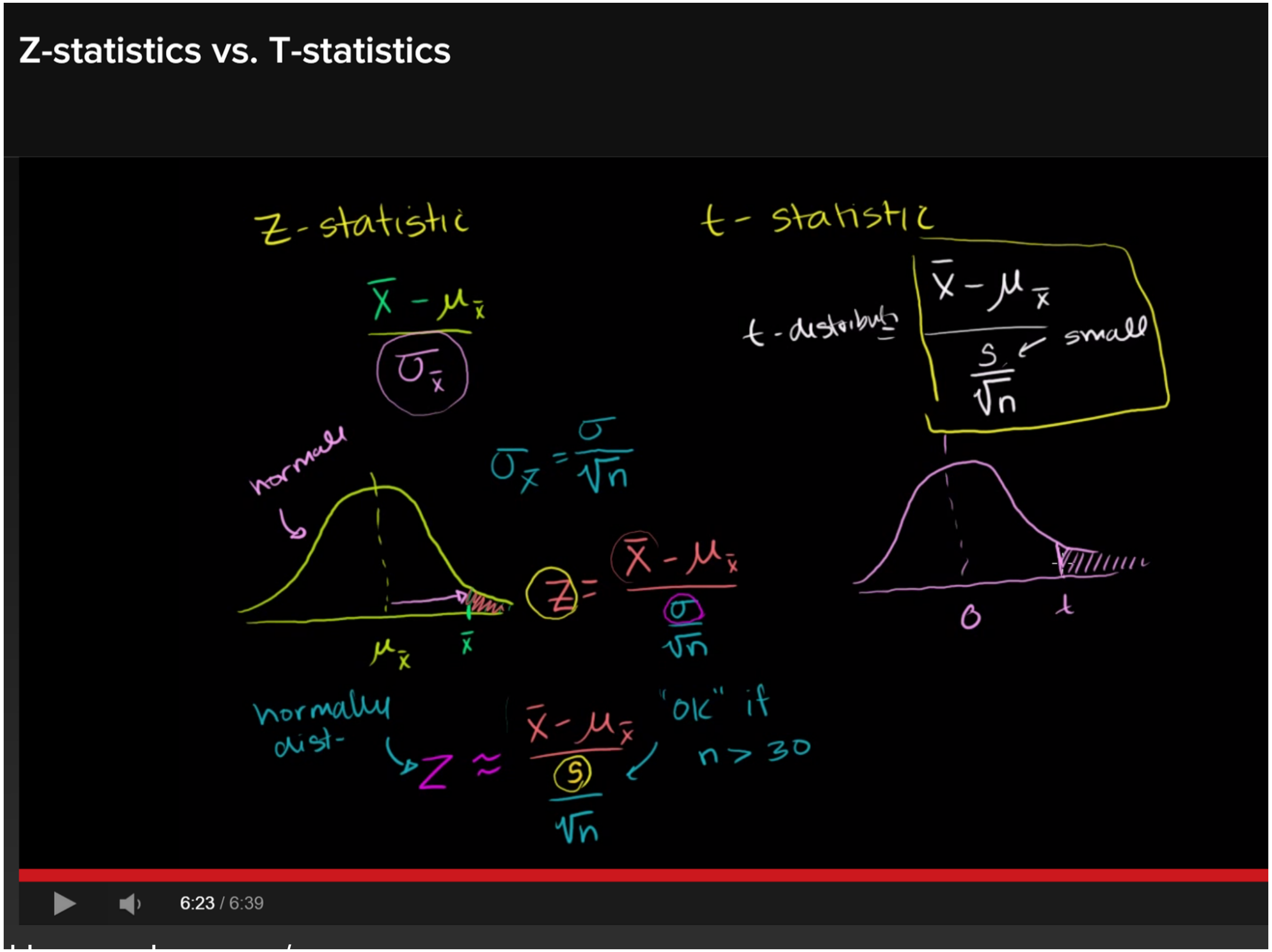
NIST/SEMATECH e-Handbook of Statistical Methods

Why are we using it? What’s an alternative?

- Quantitative understanding of data
- Exploratory Data Analysis (EDA)
  - – The primary goal of EDA is to maximize the analyst's insight into a data set and into the underlying structure of a data set
- Quantitative (Classic) Techniques
  - Hypothesis tests
  - Interval estimation
    - – An interval estimate quantifies … uncertainty in the sample estimate by computing lower and upper values of an interval

NIST/SEMATECH e-Handbook of Statistical Methods

How are we using it?



Example

- Someone claims kids at Lake Wobegon have above average intelligence
- random samples of 9 kids there with test result of {116, 128, 125, 119, 89, 99, 105, 116, 118}
- Wechsler scores (the test they take) are scaled to be normally distributed with a mean of 100 and standard deviation of 15.
- mean of sample: 112.8

http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Example

- mean(X) = 112.8
- u = 100
- test statistic: z-statistic
- H_0: u = 100
- H_α: u > 100
- $$z_{stat} = \frac{\bar{x} - \mu_0}{SEM}$$
- SEM = σ/sqrt(N) = 15/sqrt(9) = 5
- Z_{stat} = (112.8 - 100)/5 = 2.56
- look up Z table → p = 0.0052 = 0.52%

http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Example

Standard Normal Probabilities

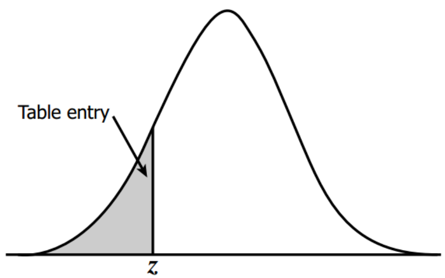
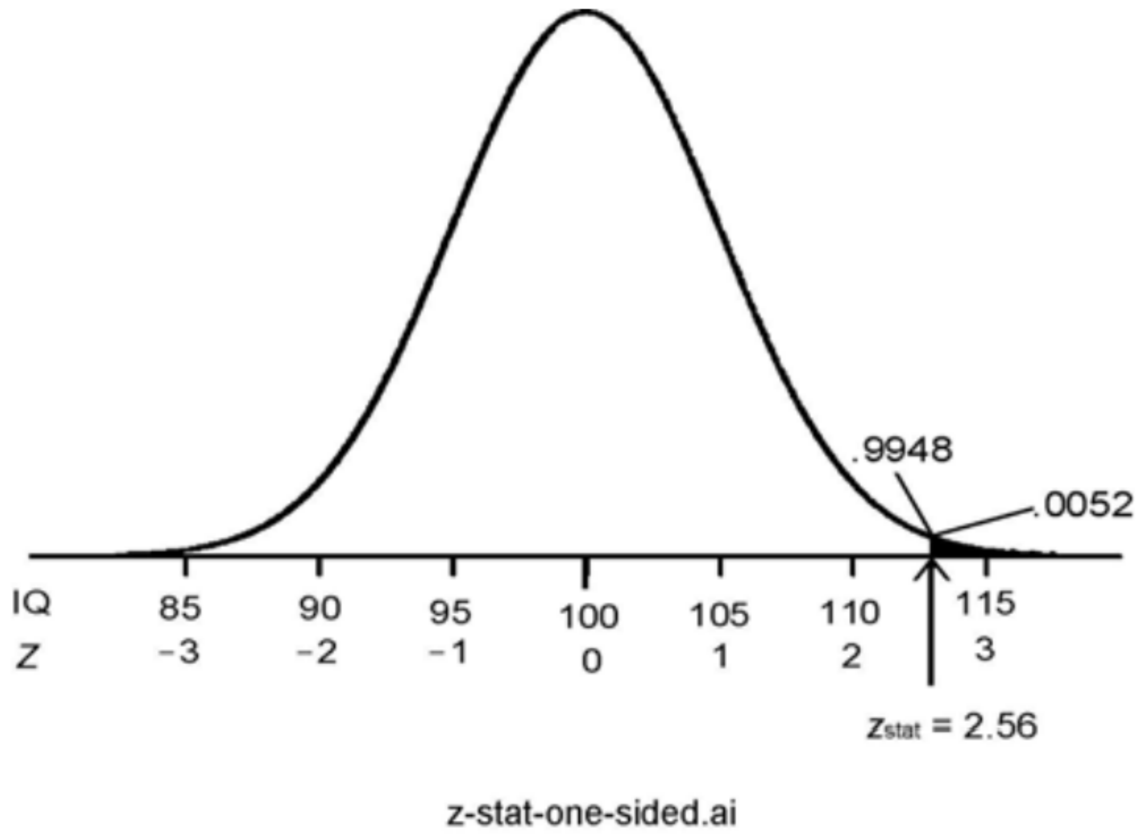


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233

http://www.stat.ufl.edu/~athienit/Tables/Ztable.pdf

Example



http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

Example

- Define conventions:
  - When p value > .10 → the observed difference is "not significant"
  - When p value ≤ .10 → the observed difference is "marginally significant"
  - When p value ≤ .05 → the observed difference is "significant"
  - When p value ≤ .01 → the observed difference is "highly significant"

http://www.sjsu.edu/faculty/gerstman/StatPrimer/hyp-test.pdf

How is it related to us?

- Test Calibration
- Test Constraints
- QA audit
- Actually test some hypothesis…

How can we use it?

- Data Acquisition:
    - Input side:
        - A sensor input value
    - Output side:
        - VSCADA
        - raw sensor data is retrievable and transformable
- Post analysis:
    - apply suitable test statistic using favorable tools