Hypothesis Testing

03/21/19

Type 1 Error ⟶ rejection of a <u>true</u> null hypothesis
"False positives"

Type 2 Error ⟶ Fail to reject <u>False</u> null hypothesis
"False Negative"

Decision About Null Hyp H_0	H_0 is True	H_0 is False
Fail to reject	correct inference $p = 1-\alpha$ "true negative"	Type II error $p = \beta$ "False negative"
Reject	Type 1 error $p = \alpha$ False positive	correct inference $p = 1-\beta$ True positive

Classical Statistical Inference (MITxOCW 6.041)

- unknown parameter ⟶ get obs x
  ↑
  just a #, not a random var (no prob dist) ⟶ Jobs to define the estimator
- probabilistic models ⟶ can deal w/ multiple hypotheses
- Estimation problems: θ unknown parameter | deterministic
  $\hat{\theta}$ is the estimate, want to estimate x └ "real #" } design an error that's not too large
- methods to do this:
  1. maximum likelihood estimation
     - highest probability / more likely to have occurred
     - vs Bayesian setting ⟶ use prior dist to compute posterior
     ⟶ Assumed prior on θ that's uniform

$$\max_{\theta} \prod_{i=1}^{n} \theta e^{-\theta x_i} \qquad x_i = \text{iid}$$

- estimator takes in random var then produces another random var
- want rv to be close to the true value of θ
  ⟶ ideally no systematic error
- Expectation w/ respect to prob dist of $\hat{\theta}$

$$E[\hat{\theta}_n] = \int f_{x_{\theta x}}(\hat{\theta}\,;\theta)\,\hat{\theta}\,d\theta$$

MSE: (from the estimator)

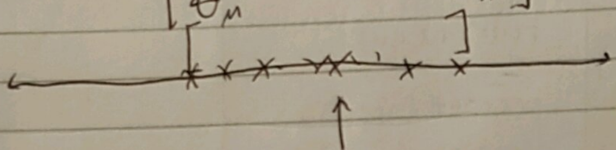
$$E[(\hat{\theta}-\theta)^2] = var[\hat{\theta}-\theta] + (E[\hat{\theta}-\theta])^2$$
$$= var(\hat{\theta}) + (bias)^2$$

$X = \hat{\theta} - \theta$

① maximum likelihood
② the method — estimate a mean (average them out)

$X_i = \theta + W_i$

$[\theta_M^-$ $\theta_M^+]$



$\hat{\theta}_M = 2.37$

95% confidence interval — "Approximately" (not exactly)

$$P\left(\hat{\Theta}_n - \frac{1.96\sigma}{\sqrt{n}} \le \theta \le \bar{\Theta}_n + \frac{1.96\sigma}{\sqrt{n}}\right) \approx 0.95$$

- if we don't know σ:
  - estimate σ from the data $\sigma = \sqrt{\theta(1-\theta)}$
  - estimating the mean of some random variable
    - plug in the estimate of the mean
    - same consistency property
- Approximate the sample mean
- Approximate the standard deviation.

Classical Inference 2

① Linear reg
① Hypothesis testing

② posterior distribution
- common mean

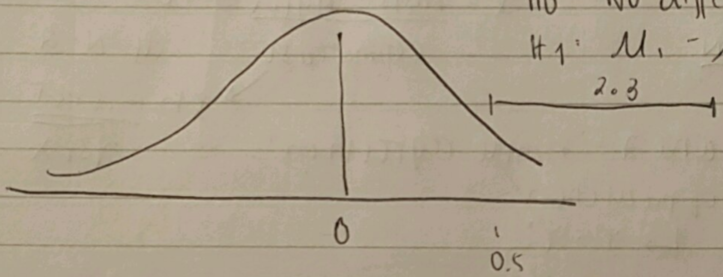
IA Section 3

03/28/19

T-Distributions

- more robust / if there's more uncertainty
- 10 df of freedom ⟶ ch 10 section 9 = correction!
- As the df increases, the t-value for 0.5 comes close to the normal standard deviation of 1.96. Check the previous chart
- inverse relationship b/n sample size?
- t ← t.test
  ↳ qt instead of t test

Confidence intervals



H_0: No difference b/n $\mu_1 - \mu_2 = 0$
H_1: $\mu_1 - \mu_2 \neq 0$

- correlation

Difference Between Means

3A Chapter 12.
- p ← c(12.7, 11.13, 10) ⟶ one tailed vs 2-tailed test

Chapter 9.
- the groups are independent
- calculating s when df are different (?)

#25 Chapter 9

Chapter 13

#3) n13 ← 16

mean = 45 − 36

sd = 10

$MSE = \frac{100 + 100}{2}$

$SE\ 13 = \sqrt{\frac{2MSE}{n}}$

Alan will not be able to convince

$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$ } for r

(b - t.05 Sb)

1

MSCA 31000 Reading Notes Week 4 08/25/19

Chapter 14 - Regression

* Need to find another resource to go over all these topics

✓ A) Introduction to Simple Linear Regression
✓ B) Partitioning sum of squares
✓ C) Standard Error of the Estimate ⟶ $\sigma_{est} = \sqrt{\frac{(1-\rho^2) SSY}{N}}$
✓ D) Inferential Statistics for b and r ⟶ Assumption: Linearity, Homoscedasticity, normal
E) Influential Observations
F) Regression Toward the mean
G) Introduction to Multiple Regression.

Significance test for the slope b:
$t = \frac{\text{stats} - \text{hypo}}{\text{estimated SE}}$ (stats ⟶ b, hypo ⟶ 0) df: N-2

$s_b = \frac{s_{est}}{\sqrt{SSX}}$

INTRODUCTION TO LINEAR REGRESSION
- y = criterion var
- x = predictor var ⟶ only one x = "simple regression"
- "best fitting line", w errors of prediction ↳ Y - Y'
  ↳ minimizes SSE
- Y' = bX + A
- computing the Regression Line:
  $b = r\frac{s_y}{s_x}$ b = slope, s = standard dev, r = correlation
  $A = M_y - bM_x$ M = mean
- standardized variables:
  M = 0, sd = 1 so: b = r, A = 0
  $z_{y'} = (r)(z_x)$ ⟵ standardized score for x; ⟶ new equation
  ↑ predicted standard score for y

Example: University GPA as a function of high school GPA

r = slope of reg for standardized values

PARTITIONING THE SUMS OF SQUARES
- Regression variation: (in y)
  ① var of predicted scores
  ② var in errors of prediction

population:
$SSY = \sum (Y - \mu_y)^2$ sum of squares

$\sum (y - M_y)^2$ sample mean M

- deviation scores vs raw scores

2

Week 4 Reading

Chapter 14 - Regression

Sum of squares of y

- deviation scores used this?
  - ↳ Deviations from the mean
- predictor variable & criterion value
- SSY

SSY'
(sum of squares predicted)

SSE
(sum of square error)

$$\boxed{SSY = SSY' + SSE}$$

Proportion explained $= \frac{SSY'}{SSY}$

Proportion not explained $= \frac{SSE}{SSY}$

$$\sigma^2_{total} = \sigma^2_{y'} + \sigma^2_e$$

Summary table

- degrees of freedom → number of predictors

Chapter 16: Transformations

- many statistical procedures work best if individual vars have certain properties
  → measurement scale & data collection

A) LOG TRANSFORMATIONS

- can be used to make highly skewed distributions less skewed
- make patterns more visible
- if the arithmetic means of 2 sets of log-transformed data are equal then the geometric mean are equal)

eg. Arithmetic mean = 1 → 3

Antilog: 10^1 → 10^3

Geometric mean $= (1 \times 10 \times 100)^{0.333} = 10$

$(3 \times 10 \times 100)^{0.333} = 1000^{1/3}$

B) TUKEY LADDER OF POWERS

- Bivariate data → Tukey re-expressed vars using power transformation

① Plot on scatter gram

- polynomial regression

λ = parameter chosen to make the relationship as close to a straight line as possible

$y = b_0 + b_1 x^\lambda$ or $y^\lambda = b_0 + b_1 x$

$\lambda = 1$ → data is unchanged

redefine Tukey's transformation to be $-(x^\lambda)$ if $\lambda < 0$

formally defined as:

$$\bar{x}_\lambda = \begin{cases} x^\lambda & \text{if } \lambda > 0 \\ \log x & \lambda = 0 \\ -(x^\lambda) & \lambda < 0 \end{cases}$$

- the Best Transformation for linearity

- the goal is to find a value λ that makes the scatter diagram as linear as possible
- r is a measure of the linearity of a scatter diagram

Reducing skew

- Decreasing λ makes the distribution less positively skewed

Fundamentals of Statistics (MIT OCW 18.6501x (Fall 2016)

- Professor Philippe Rigollet ⟶ MIT
- unit 4 - Hypothesis Testing
- unit 6 - Linear Regression
- unit 2 - Inference

I. Probability and Linear Algebra Review

- Gaussian Probabilities
- What is statistics?
  - mathematical theory behind statistical models
  - How to lie w/ statistics
- Probability

Lecture Reading Notes (weeks) 04/01/19

①

CHAPTER 15. ANALYSIS OF VARIANCE

INTRO.
- statistical method used to test the difference between 2 or more means
  - ⟶ inferences about the mean can be made by analyzing the variance
  - ⟶ used to test general rather than specific differences among means
- ANOVA tests the non-specific null hypothesis that all population means are equal, eg

$$\mu_{false} = \mu_{felt} = \mu_{miserable} = \mu_{neutral}$$

⟶ Also called the OMNIBUS NULL HYPOTHESIS

- Rejecting the omnibus null hypothesis means that at least one population mean is different from at least one other mean.
- But we don't know which mean is different from which, so Tukey HSD test more specific