Ch. 9 INFERENTIAL STATS

SYMBOLS

	PARAMETRIC (POPULATION)	STATISTIC (SAMPLE)
MEAN	μ, μ_1, μ_2	$\bar{x}$, $\bar{x}_1$, $\bar{x}_2$
VARIANCE	σ^2, σ_1^2, σ_2^2	s^2, s_1^2, s_2^2
SD	σ, σ_1, σ_2	s, s_1, s_2
PROPORTION	p, p_1, p_2	$\hat{p}$, $\hat{p}_1$, $\hat{p}_2$
SIZE	(none)	n, n_1, n_2

9.1 POINT-ESTIMATE

- a single number. use a number computed from a sample to estimate the parameter

THE TWO METHODS

METHOD OF MOMENT

FIND MOMENT OF POPULATION → $M_1 = E(X)$, $M_2 = E(X^2)$ (POP. MEAN, VARIANCE)

MOMENT OF SAMPLE → $m_1 = \frac{1}{n}\sum X_i = \bar{x}$ mean

$m_2 = \frac{1}{n}\sum X_i^2$

ALL INDEPENDENT N IDENTICAL DISTRIBUTION

SET THEM EQUAL.

GET EQUATION + SOLVE

METHOD OF MAXIMUM LIKELYHOOD

FIND JOINT PROBABILITY MASS FUNCTION (PMF)

OR JOINT PROBABILITY DENSITY FUNCTION (PDF)

FOR $x = (x_1, x_2, x_3, \dots x_n)$

MAXIMIZE IT BY TAKING DERIVATIVES

SET DERIVATIVE = 0

SIDENOTE 4 VARIANCE:

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ in probability chapter 8 ← UNBIASED } referred to as sample variance

$m_2' = \frac{\sum (x_i - \bar{x})^2}{n}$ in THIS chapter ← BIASED } referred to as sample central moment

LOG RECAP

$\ln e^{\square} = \square$

$\ln x^n y^m = \ln x^n + \ln y^m$

$= n \ln x + m \ln y$

$\frac{d}{d\triangle} \ln \triangle = \frac{1}{\triangle}$

CALC RECAP

$\frac{d}{dx} \ln f(x) = 0$ → x = CRITICAL #

EXAMPLE FOR DISCRETE VARIABLE

X IS A D.V.

x	3	7
P(X)	θ	$1-\theta$

$(X_1, X_2, \dots, X_8) = (7, 3, 3, 7, 3, 7, 7, 7)$

THESE ARE RANDOM DATA SAMPLE

USE THEM TO:

θ = parameter

GOAL: ESTIMATE θ

· M. O. MOMENT

$E(X) = \sum X_i P(X_i)$

1) 1ST POP. MOMENT (M_1) : $E(X) = 3\theta + 7(1-\theta) = 7 - 4\theta$

2) 1ST SAMPLE MOMENT $= \frac{1}{n}\sum X_i = \bar{x}$

$\bar{x} = \frac{7+3+3+7+3+7+7+7}{8} = 5.5$

SET = TO $\bar{x}$

$7 - 4\theta = \bar{x}$ → $\hat{\theta} = \frac{7 - \bar{x}}{4}$ } GENERAL FORMULA

HAT BECAUSE ITS AN ESTIMATE

$= 0.375$

M.O. MAX. LIKELIHOOD

each are independent + identical

$f(x_1, x_2, x_3, \dots, x_8) = f(7, 3, 3, 7, 3, 7, 7, 7)$

JOINT PMF $= (1-\theta)^5 \theta^3$ (7's, 3's)

1) let $f = \theta^3 (1-\theta)^5$

2) take derivative → $\ln f = 3 \ln \theta + 5 \ln (1-\theta)$

$\frac{d}{d\theta} \ln f = \frac{3}{\theta} - \frac{5}{1-\theta}$

3) set = 0

$\frac{3}{\theta} - \frac{5}{1-\theta} = 0$, $\theta = \frac{3}{8}$

$\hat{\theta} = \frac{3}{8} = 0.375$

→ NOT GUARANTEED BOTH METHOD SAME ANSWER.

EXAMPLE FOR CONTINUOUS VARIABLE

X HAS A NORMAL DISTRIBUTION W/ UNKNOWN μ + σ

$X \sim$ normal $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

$X_1, X_2, X_3, \ldots, X_n \sim f(x)$ ← TREAT X_i as individual fixed numbers.

↳ USE THEM TO FIND MEAN + σ

WE WANT JOINT PDF $f(X_1, X_2, \ldots X_n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(X_1-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(X_2-\mu)^2}{2\sigma^2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(X_n-\mu)^2}{2\sigma^2}}$

JOINT PDF = $f = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2}$

take derivative of $\ln f$

M.O. LIKELIHOOD

$\ln f = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \Sigma (x_i-\mu)^2$

$\mu + \sigma$ are parameters, we find them

we take partial derivative

$\frac{d}{d\mu} \ln f = \frac{1}{2\sigma^2} \Sigma\, 2(x_i-\mu) = \frac{1}{\sigma^2}(\Sigma x_i - n\mu) = 0$

$\mu = \frac{\Sigma x_i}{n} = \bar{x}$, $\mu = \bar{x}$

$\ln f = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \Sigma(x_i-\mu)^2$

$\frac{d}{d\sigma} \ln(f) = 0 - \frac{n}{\sigma} + \frac{1}{\sigma^3} \Sigma(x_i-\mu)^2 = 0$

$\sigma^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$, $\boxed{\hat{\sigma}^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}}$

M.O. MOMENT

first pop. moment = $E(X) = \mu$
first sample moment = $\frac{\Sigma x_i}{n} = \bar{x}$
} set =

$\mu = \bar{x}$, $\hat{\mu} = \bar{x}$

second pop. central moment = $E(X-\mu)^2 = \sigma^2$
second sample central moment = $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$
} set =

$\sigma^2 = \frac{1}{n}\Sigma(x_i-\bar{x})^2$, $\hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i-\bar{x})^2$

9.2 INTERVAL ESTIMATE

// MORE PRACTICAL

WE USE SAMPLE TO COMPUTE INTERVAL

ANSWER IS ALWAYS IN INTERVAL: EX - (2,3), $(-\infty, 4]$, $[-1.6, \infty)$

AN INTERVAL IS CORRECT IF IT CAPTURES THE PARAMETER

HOW IT'S COMPUTED → $\hat{\theta} \pm ME$

POINT ESTIMATE

MARGIN OF ERROR

RELIABILITY

↳ THE CONFIDENCE % OF GETTING THE CORRECT PARAMETER

↳ USUALLY 90% - 95%

PRECISION

WIDTH OF INTERVAL = 2ME

↳ WIDTH OF THE INTERVAL

↳ THE NARROWER PROVIDES MORE INFO

MORE RELIABLE = LESS PRECISE
MORE PRECISE = LESS RELIABLE

THE TRADE OFF

TO GET ↑R + ↑P, NEED LARGER SAMPLE SIZE

HIGH RETURN, HIGH RISK

MORE COSTLY

ONE MEAN μ

LARGE SAMPLE

$n \geq 30$

CALCULATOR

STAT > TEST

(L) 1-SAMPLE Z-INTERVAL

(S) 1-SAMPLE T-INTERVAL

Z-INTERVAL: $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

CRITICAL VALUE

we use this formula bc CLT

SMALL SAMPLE

$n < 30$

+ NORMAL DISTRIBUTION

T-INTERVAL: $\bar{x} \mp t_{\alpha/2} \frac{s}{\sqrt{n}}$, $df = 1$

(IF NOT NORMAL, WE DONT DO IT)

ONE PROPORTION p

CALCULATOR

STAT > TEST

1-PORP Z-INTERVAL

LARGE SAMPLE

$n\hat{p} \geq 10,\ n(1-\hat{p}) \geq 10$

Z INTERVAL: $\hat{p} \mp z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

HAT ALWAYS FROM SAMPLE
THIS IS THE PT. ESTIMATE PORPORTION

also based on CLT, so REQUIRE LARGE SAMPLE

SMALL SAMPLE

WE DONT DO

SIDES

$2 < \mu < 3$ TWO SIDED WE ONLY USE THIS ONE

$\mu < 3$ ONE SIDED

WHEN TO USE WHAT?

#1) FIND KEYWORDS

↳ INTERVAL, POINT, HYPOTHESIS

↳ MEAN, PERCENTAGE, PROPORTION

#2) ONE OR TWO VARIABLES INVOLVED?

#3) LARGE OR SMALL SAMPLE?

TWO MEANS $\mu_1 - \mu_2$

DIFFERENCE OF 2 POP. MEANS

(TWO SAMPLES ARE INDEPENDENT)

LARGE SAMPLE

both $n \geq 30$

Z-INTERVAL: $\bar{x}_1 - \bar{x}_2 \mp z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

comparing 2 populations

WE COMPUTE ↳ ONE INTERVAL

$1 < \mu_1 - \mu_2 < 3$

↳ $\mu_1 - \mu_2 > 0$

$= \mu_1 > \mu_2$

↳ $\mu_1 - \mu_2 < 0$

$= \mu_1 < \mu_2$

$-1 < \mu_1 - \mu_2 < 1$

↳ $\mu_1 - \mu_2 = 0$

$= \mu_1 = \mu_2$

THIS TELLS YOU NOTHING THO, NO CONCLUSION CAN BE DRAWN

IF ONE LARGE, ONE SMALL ↳ USE SMALL!

SMALL SAMPLE

NORMAL DISTRIBUTION REQUIRED FOR BOTH POPULATIONS

POOLED IF $\sigma_1 = \sigma_2$

$\bar{x}_1 - \bar{x}_2 \mp t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$S_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$

POOLED: PUTTING THE 2 SAMPLES TOGETHER SINCE THEY HAVE THE SAME SD.

WITH $df = n_1 + n_2 - 2$

UNPOOLED $\sigma_1 \neq \sigma_2$

$\bar{x}_1 - \bar{x}_2 \mp t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

WE USUALLY USE THIS BC WE DONT KNOW σ_1 OR σ_2

2 POPULATION PERCENTAGE $P_1 - P_2$

LARGE SAMPLE

$n\hat{p} \geq 10$, $n(1-\hat{p}) \geq 10$ for both samples

2 PROP Z TEST : $\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

9.3 HYPOTHESIS TEST

THE TYPE OF PROBLEM WHERE WE HAVE A NULL HYPOTHESIS + AN ALTERNATE HYPOTHESIS

ASSUMED AT FIRST

EX: PERON ON TRIAL IS INNOCENT

NEED TO PROVE W/ EVIDENCE

EX: PERSON IS PROVEN GUILTY

WE FIND ENOUGH EVIDENCE TO PROVE NULL HYPOTHESIS IS FALSE

CANNOT FIND ENOUGH EVIDENCE TO REJECT NULL ≠ FIND EVIDENCE TO SHOW NULL IS TRUE

IF WE CANT FIND ENOUGH EVIDENCE TO PROVE THE ALT. HYPOTHESIS, THEN NULL IS TRUE

H_0 : NULL HYPOTHESIS

H_a : ALTERNATE HYPOTHESIS

TWO TAILED	RIGHT TAILED	LEFT TAILED
$H_0 : \mu_1 - \mu_2 = 0$	$H_0 : \mu_1 - \mu_2 \leq 0$	$H_0 : \mu_1 - \mu_2 \geq 0$
$H_a : \mu_1 - \mu_2 \neq 0$	$H_a : \mu_1 - \mu_2 > 0$	$H_a : \mu_1 - \mu_2 < 0$

!! THE = ALWAYS !!

!! ON H_0 !!