ProbStats L15a

Hypothesis Testing

Tea Tasting (Fisher's Exact Test)

April 6, 2023

Ronald Fisher

Lady tea + milk : tea first
or
milk first

8 cups : 4 cups tea first
4 cups milk first

Two views

⓪ same → random guess — Null Hypothesis

① she can tell the difference.

Lady Guess

Truth

	Milk First	Tea First
Milk First	3 = k	1
Tea First	1	3

k = 2

2 = k	2
2	2

k = 4

4	0
0	4

"all correct"

0	4
4	0

k = 0

k = 1

1	3
3	1

$$Pr(\text{"all correct"}) = \frac{1}{\text{"number of ways to guess"}} = \frac{1}{\binom{8}{4}} = \frac{1}{70} \approx 0.014$$

		Lady	
		Milk	Tea
Truth	Milk	k	$4-k$
	Tea	$4-k$	k

$k = \{0, 1, 2, 3, 4\}$

$$P(k) = \frac{\binom{4}{k}\binom{4}{4-k}}{\binom{8}{4}} = \frac{1}{70} \cdot \binom{4}{k}^2$$

hypergeometric distribution

k	0	1	2	3	4
$P(k)$	$1/70$	$16/70$	$36/70$	$16/70$	$1/70$
$\Pr(K \geq k)$	1	$69/70$	$53/70$	$17/70$	$1/70$

$\Pr(k \geq 3) \approx 0.24$

$\Pr(k \geq 4) \approx 0.014$

Goal: get high score

score ↳ k

k	0	1	2	3	4	5	6
$P(k)$	2/10	1/20	2/10	2/10	2/10	1/20	1/10
$\Pr(X \geq k)$	1	16/20	15/20	11/20	7/20	3/20	1/10

$\Pr[k=5] = 1/20$

$\Pr[k=6] = 1/10$

$\Pr[k \geq 5]$

Summary of Hypothesis Test

1. Define Null Hypothesis

2. Assuming Null Hypo is true
   → determine probability of outcomes

3. Collect Data

4. Compute Probability of data outcome or something more extreme.

Quiz Review

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

param

Sample $X_1 \dots X_n \overset{\sim}{iid} f(\theta) = N(\mu, \sigma^2)$

estimate μ

statistic $T(X_1 \dots X_n)$ R.V.

eg. $\alpha = 0.05 \Rightarrow 95\%$

$(1-\alpha)100\%$ - confidence interval

$[L_n, R_n]$ so $\Pr(L_n \le \mu \le R_n) = 1-\alpha$

$L_n = \bar{X}_n - \underbrace{z_{\alpha/2}}\frac{\sigma}{\sqrt{n}}$ if we know σ^2

$z_{\alpha/2} = 1-\alpha/2$ - quantile $N(0,1)$

$\bar{X}_n - t_{\alpha/2}\frac{S_n}{\sqrt{n}}$ if we do <u>not</u> know σ^2 t-dist

$t_{\alpha/2} = 1-\alpha/2$ - quant t

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$ Sample Variance

$$\sqrt{S_n^2} = S_n$$ = sample std. dev.

---

t-distribution $X_i \sim t(n-1)$

↑ degrees of freedom

$z_{\alpha/2}$ = $1-\alpha$ quantile of $\mathcal{N}(0,1)$

$t_{\alpha/2}$ = $1-\alpha$ quantile of t-distribution $\to t(n-1)$

pdf f_{μ,σ^2} from $N(\mu,\sigma^2$

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$E[X] = \mu$

$X \sim N(\mu,\sigma^2)$

$Var[X] = \sigma^2$