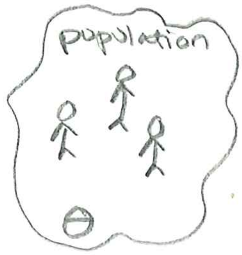
Ex. 1 Broccoli

θ: mean consumption of Broccoli by Norwegian in one year



Research question:

$H_0: \theta \le 4\text{kg}$ $H_1: \theta > 4\text{ kg}$

$(\theta = 4\text{kg})$

Method:

Sample n persons randomly, observe (ask) broccoli consumption

$X_1, X_2, \dots, X_n$,

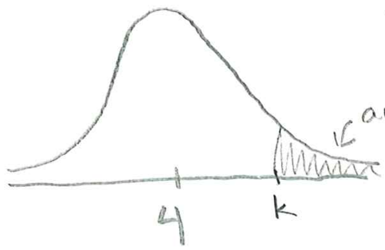
calculate mean $\bar{X}$.

What can occur "by chance" (H_0)? (1)

Distn. of $\bar{X}$:

$\bar{X} \overset{H_0}{\sim} N(4, \sigma^2/n)$

↑ assume $\frac{\sigma^2}{n} = 1$



What is evidence of H_1?

(A) $\bar{X} > 4$ points to H_1 (slide table)

but want $P(\text{type I error}) = \alpha$

↳ $\bar{X} > k$ evidence of H_1

$P(\bar{X} > k \mid H_0) = \alpha$

$\alpha = 0.05$
$k = 5.64$

Perform study

$\bar{x} = 4.22\text{ kg}$

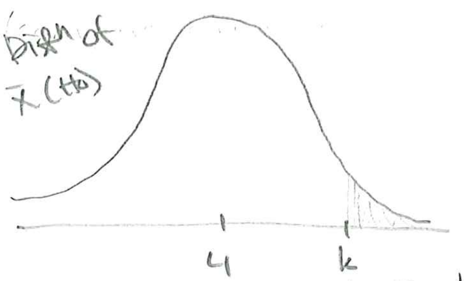
compare $\bar{x}$ to distn under H_0

↳ Not reject.

What is evidence of H_1?

(B) p-value $w(\bar{x}) < \alpha$ is evidence of H_0

$W(\bar{x}) = P(\bar{X} > \bar{x} \mid H_0)$



all observations here give p-value $< \alpha$ (don't need to know k)

Small p-values: our data is unlikely to come from H_0 (by chance)

Perform study:

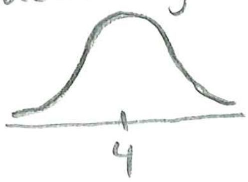
$\bar{x} = 4.22\text{kg}$

$w(\bar{x}) = 0.41$

↳ compare to sign. level α

↳ not reject.

If H_0 true and we repeat experiment many times, $\bar{X}$ takes values according to



but what about $w(\bar{X})$, the p-values?

Hint: $w(\bar{x})$ is a cumulative probability

[R example]

Ex. 1 cont.

Here CDF $N(4,1)$-distn.

observation $\bar{x}$:

$$W(\bar{x}) = P(\bar{X} > \bar{x} \mid H_0) = 1 - F(\bar{x})$$

p.v. $\bar{X}$:

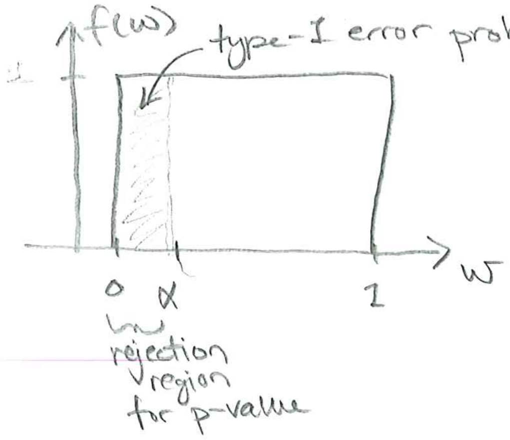
$$W(\bar{X}) = 1 - F(\bar{X})$$

$$P(W(\bar{X}) \leq w) = P(1 - F(\bar{X}) \leq w) = P(F(\bar{X}) \geq 1-w)$$

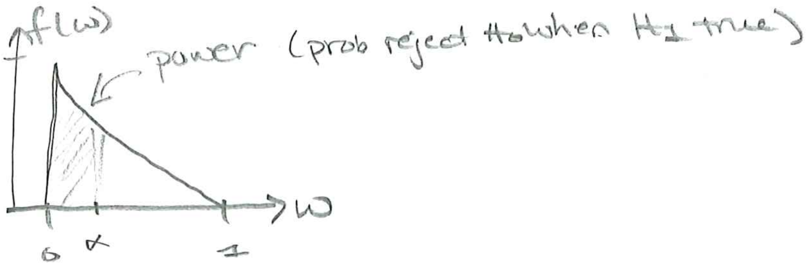
$$= 1 - P(F(\bar{X}) \leq 1-w) = 1 - P(\bar{X} \leq F^{-1}(1-w)) \qquad [P(\bar{X} \leq a) = F(a)]$$

$$= 1 - F(F^{-1}(1-w)) = 1 - (1-w) = w$$

$\Rightarrow W(\bar{X}) \sim \text{Uniform}(0,1)$ [H_0 true]



H_1 true : small p-values occur much more often



Note : Exact p-value (as in example 1)

$$P(W \leq \alpha \mid H_0) = \alpha$$

Valid p-value

$$P(W \leq \alpha \mid H_0) \leq \alpha \qquad \text{(diskret test. obs)}$$

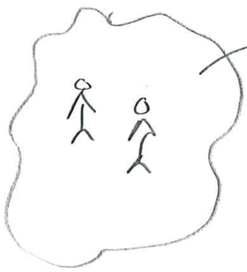
Ex. 2 Fruits and veg.

θ_1 : Broccoli $\qquad H_1 : \theta_1 > 4kg$

θ_2 : Apple $\qquad H_1 : \theta_2 > 8kg$

etc.

5 tests.



sample

$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_5$

$w_1, w_2, \dots, w_5 \leftarrow$ 5 p-values (uniform(0,1) if H_0-s are true)

Should all p-values be compared to 0.05?

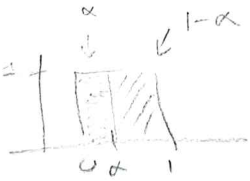
Note : Ex 1 : $P(\text{type-1 error}) = P(W_1 < \alpha \mid H_0) = \alpha$.
false positive

Now 5 tests

Reject H_0 if $w_j < \alpha$ for $j = 1, \dots, 5$

Assume $W_1, \dots, W_5$ independent RV,

assume all H_0-s true.



$$P(\text{at least one false pos}) = 1 - P(\text{no false pos.}) \overset{\text{ind}}{=} 1 - \prod_{j=1}^{5} \underbrace{P(W_j > \alpha \mid H_0)}_{\text{not reject}}$$

$$\underset{H_0\text{s true, uniform(0,1)}}{=} 1 - \prod_{j=1}^{5} (1-\alpha) = 1 - (1-\alpha)^5$$

$\alpha = 0.05 \longrightarrow P(\text{at least one false pos}) = \underline{0.23}$.

Def : Family-wise error rate FWER

Prob of at least one false positive when we perform multiple tests (all H_0 true)

Aim: control FWER at some level α.

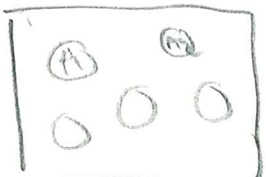
Method 1 : Bonferroni

We want $\quad P(\underbrace{W_1 < \alpha_{loc}}_{A_1} \cup \underbrace{W_2 < \alpha_{loc}}_{A_2} \cup \dots \cup \underbrace{W_m < \alpha_{loc}}_{A_m}) = \alpha \quad (H_0)$

union = total area. diff to find.

Boole's inequality

$$P(A_1 \cup A_2 \cup \dots \cup A_m) \leq \sum_{j=1}^{m} P(A_j)$$



Bonf. cont.

exact p-values, Uniform(0,1)

$$P(W_1 < \alpha_{loc}, \ldots, W_m < \alpha_{loc}) \overset{\text{Boole}}{\leq} \sum_{j=1}^{m} P(W_j < \alpha_{loc}) = \sum_{j=1}^{m} \alpha_{loc}$$

$$= m\alpha_{loc} = \underset{\uparrow \text{FWER}}{\alpha}$$

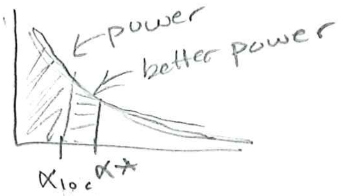
$$\boxed{\text{set } \alpha_{loc} = \frac{\alpha}{m}}$$

EX 2 cont. $m = 5$, $\alpha_{loc} = \frac{0.05}{5} = 0.01$

controls FWER at 5%.

[If time: the problem with conservative ← due to Boole.

H_1 true



]

Method 2: Šidák ← more assumptions, less conservative

- Assume independent p-values $W_1, W_2, \ldots, W_m$

$$P(\text{at least one false pos}) = 1 - P(\text{no false pos}) = 1 - \prod_{j=1}^{m} P(W_j > \alpha_{loc} \mid H_0)$$

$$\overset{\text{uniform}(0,1)}{=} 1 - \prod_{j=1}^{m} (1 - \alpha_{loc}) = 1 - (1 - \alpha_{loc})^m = \alpha \quad \text{(Solve for } \alpha_{loc}\text{)}$$

$$\rightarrow \boxed{\text{Set } \alpha_{loc} = 1 - (1-\alpha)^{1/m}}$$

Ex. 2 cont, $m = 5$, $\alpha = 0.05 \rightarrow \alpha_{loc} = 0.0102$

Indepence assumption ok?

Exam 2024

FWER $\alpha = 0.01$, $m = 3$

Bonferroni $\alpha_{loc} = \frac{0.01}{3} = 0.0333$

Sidak $\alpha_{loc} = 1 - (1 - 0.01)^{1/3} = 0.00334$

reject $H_0: \beta_2 = 0$ (third test)

Exam 2023

FWER $\alpha = 0.05$, $m = 3$

Bonferroni $\alpha_{loc} = \frac{0.05}{3} = 0.0167$

reject two first hyp.