# Hypothesis testing - a review

## Statistical Learning - Modern Inference

Aldo Solari

# Darwin data

- Charles Darwin collected data on Zea mays plants
- The plants were descended from the same parents and planted at the same time.
- Half of the plants were *self-fertilized*, and half were *cross-fertilized*
- Planted *in pairs* in different pots
- Purpose of the experiment: compare *heights*

| | Pot | Cross | Self |
|---|---|---|---|
| 1 | I | 23.500 | 17.375 |
| 2 | I | 12.000 | 20.375 |
| 3 | I | 21.000 | 20.000 |
| 4 | II | 22.000 | 20.000 |
| 5 | II | 19.125 | 18.375 |
| 6 | II | 21.500 | 18.625 |
| 7 | III | 22.125 | 18.625 |
| 8 | III | 20.375 | 15.250 |
| 9 | III | 18.250 | 16.500 |
| 10 | III | 21.625 | 18.000 |
| 11 | III | 23.250 | 16.250 |
| 12 | IV | 21.000 | 18.000 |
| 13 | IV | 22.125 | 12.750 |
| 14 | IV | 23.000 | 15.500 |
| 15 | IV | 12.000 | 18.000 |

# Questions

| | type | average height |
|---|---|---|
| 1 | Cross | 20.19 |
| 2 | Self | 17.57 |

1. Is the difference in heights too large to have occurred by chance?
2. Can we estimate the height increase, and assess the uncertainty of our estimate?

# Hypothesis testing

$H_0$ : There is no difference in height between cross-fertilized and self-fertilized plants

Hypothesis testing is a type of *stochastic proof by contradiction*

## Deterministic proof by contradiction

1. Assume a proposition, the opposite of what you think about, i.e. the opposite conclusion of your theorem
2. Write down a sequence of logical steps/math
3. Derive a contradiction
4. Conclude that the proposition is false (which implies that the theorem is true)

## Stochastic proof by contradiction

1. Set $H_0$ (the proposition)
2. Collect data (which is noisy)
3. Derive an apparent contradiction (i.e. if $H_0$ is true, then this data is very weird)
4. Hence we reject $H_0$; this is called a "discovery"

Hypothesis testing is stochastic because we might make errors: *Type I* (false discoveries) and *Type II* (missed discoveries)

# Galton model

Galton considered a model where the height of a self-fertilized plant is

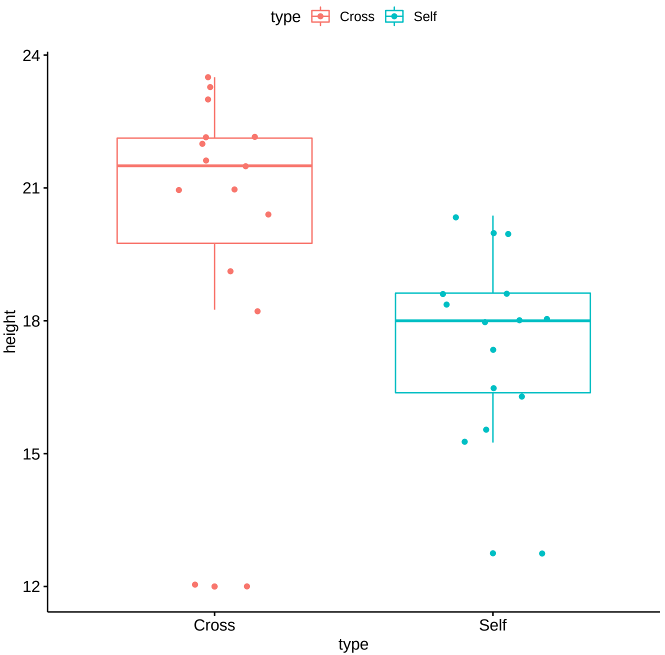
$$Y = \mu + \sigma\varepsilon$$

and of a cross-fertilized plant is

$$X = \mu + \theta + \sigma\epsilon$$

where $\mu$, $\theta$ and $\sigma$ are unknown parameters, and $\varepsilon$ and $\epsilon$ are independent random variables with mean 0 and variance 1

Self-fertilized plants: $Y_1, \ldots, Y_{15}$ i.i.d. as $Y$
Cross-fertilized plants: $X_1, \ldots, X_{15}$ i.i.d. as $X$.


type
Cross
Self
24
21
18
15
12
height
Cross
Self
type

# Two-sample t-test

1. Is the average height increase $\theta \neq 0$?
2. Can we estimate $\theta$, and assess the uncertainty of our estimate?

If we assume that $\varepsilon$ and $\epsilon$ have a $N(0, 1)$ distribution, we can use

```
Two Sample t-test


data:  height by type
t = 2.4371, df = 28, p-value = 0.02141
alternative hypothesis: true difference in means is not equ
95 percent confidence interval:
 0.4173433 4.8159900
sample estimates:
mean in group Cross  mean in group Self
           20.19167            17.57500
```

# Fisher model

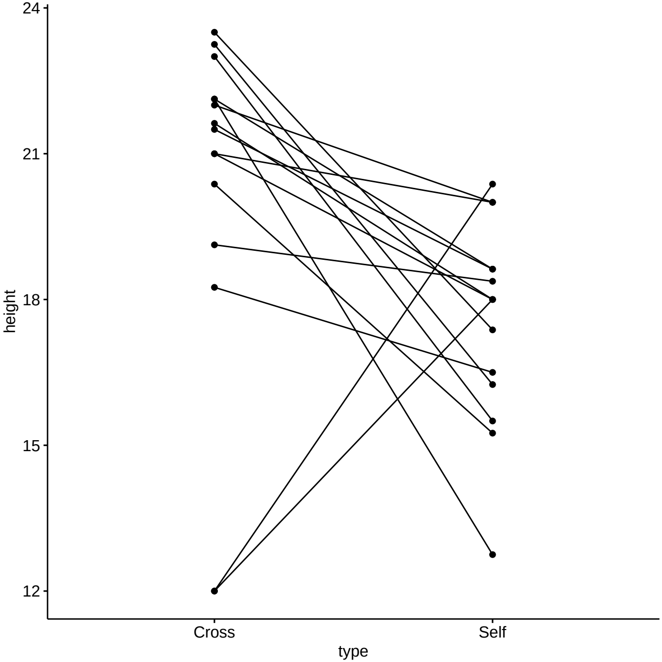
Comparison of different pairs would involve differences in humidity, growing conditions, and lighting, which are not of interest, whereas *comparison within pairs* would depend only on the type of fertilization. Fisher considered the model

$$Y_i = \mu_i + \sigma\varepsilon_i, \quad X_i = \mu_i + \theta + \sigma\epsilon_i, \quad i = 1, \dots, n$$

The parameter $\mu_i$ represents the effects of the planting conditions for the $i$th pair, and $\varepsilon_i$ and $\epsilon_i$ are independent random variables with mean 0 and variance 1.
The $\mu_i$ could be eliminated by using the differences

$$D_i = X_i - Y_i$$

which have mean $\theta$ and variance $2\sigma^2$.


24
21
18
15
12
height
Cross
Self
type

# Paired t-test

If we assume that $\varepsilon$ and $\epsilon$ have a $N(0, 1)$ distribution, we can use

```
One Sample t-test

data:  differences
t = 2.148, df = 14, p-value = 0.0497
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.003899165 5.229434169
sample estimates:
mean of x
 2.616667
```

# P-values

- Choose a test statistic $T = t(Y)$, large values of which cast doubts on $H_0$
- Observe the data $y_{obs}$, realization of $Y$
- $$p_{\mathrm{obs}} = \mathrm{P}_0(T \geq t_{\mathrm{obs}})$$

  where $t_{\mathrm{obs}} = t(y_{obs})$ and $\mathrm{P}_0$ is the probability under $H_0$

# P-value null distribution

- $p_{\mathrm{obs}} = 1 - F_0(t_{\mathrm{obs}})$, where $F_0$ is the null distribution function of $T$, supposed to be continuous and invertible.
- One interpretation of $p_{\mathrm{obs}}$ stems from the corresponding random variable $P = 1 - F_0(T)$
- The null distribution of $P$ is *Uniform*(0,1): for any $u \in (0,1)$,

$$\mathrm{P}_0(P \leq u) = \mathrm{P}_0(F_0^{-1}(1-u) \leq T) = 1 - F_0(F_0^{-1}(1-u)) = u$$

# Valid p-values

We have a *valid test* if its $p$-value is uniformly distributed under $H_0$, i.e.

$$\mathrm{P}_0(P \leq u) = u \quad \forall u \in (0,1) \tag{1}$$

or more generally if the $p$-value is *stochastically dominated* by the uniform distribution under $H_0$, i.e.

$$\mathrm{P}_0(P \leq u) \leq u \quad \forall u \in (0,1) \tag{2}$$

# One- and two-sided tests

- Suppose that we have a test statistic $T$ (with continuous distribution), small and large values of which indicate a departure from $H_0$
- Calculate

$$p_{\text{obs}}^{-} = \mathrm{P}_0(T \leq t_{\text{obs}}), \quad p_{\text{obs}}^{+} = \mathrm{P}_0(T \geq t_{\text{obs}})$$

- The $p$-value is

$$p_{\text{obs}} = 2\min(p_{\text{obs}}^{-}, p_{\text{obs}}^{+})$$

  This follows because the null distribution of $Q = \min(P^{-}, P^{+})$ is

$$Q = \min(1 - U(0,1), U(0,1)) = U(0,1/2)$$

  thus the null distribution of $2Q$ is $U(0,1)$

# Nonparametric tests

Consider a more general matched pair model for the Darwin data:

$$Y_i = \mu_i + \sigma_i \varepsilon_i, \quad X_i = \mu_i + \theta + \tau_i \epsilon_i, \quad i = 1, \ldots, n$$

The height differences may be written as

$$D_i = \theta + (\tau_i \epsilon_i - \sigma_i \varepsilon_i)$$

If we assume that $\varepsilon_i$ and $\epsilon_i$ are independent and symmetrically distributed around 0, then $D_i$ is symmetrically distributed around $\theta$, i.e.

$$D_i - \theta \stackrel{d}{=} \theta - D_i \qquad i = 1, \ldots, n$$

# Sign test

- If $H_0 : \theta = 0$ is true, then the probability that $D_i$ falls on either side of 0 is $1/2$ (because $D_i$ is symmetric around 0)
- $$T = \sum_{i=1}^{n} \mathbb{1}\{D_i > 0\}$$
- Under $H_0$, $T \sim \mathrm{Binomial}(n, 1/2)$
- $$p_{\mathrm{obs}}^{+} = \mathrm{P}_0(T \geq t_{\mathrm{obs}}) = \sum_{k=t_{\mathrm{obs}}}^{n} \binom{n}{k} \frac{1}{2^n}$$
  $$p_{\mathrm{obs}}^{-} = \mathrm{P}_0(T \leq t_{\mathrm{obs}}) = \sum_{k=0}^{t_{\mathrm{obs}}} \binom{n}{k} \frac{1}{2^n}$$

  In a discrete problem $p_{\mathrm{obs}}$ is $q_{\mathrm{obs}} = \min(p_{\mathrm{obs}}^{-}, p_{\mathrm{obs}}^{+})$ plus the achievable $p$-value from the other tail of the distribution nearest to but not exceeding $q_{\mathrm{obs}}$.

```
Exact binomial test

data:  13 and 15
number of successes = 13, number of trials = 15,
p-value = 0.007385
alternative hypothesis:
true probability of success is not equal to 0.5
95 percent confidence interval:
 0.5953973 0.9834241
sample estimates:
probability of success
             0.8666667
```