Two sample inference

Hypothesis tests and intervals for comparing means from paired and independent data

Today’s agenda

1. Reading quiz [2pm section] [4pm section]
2. [lecture/lab] Two-sample inference for population means
   a. Paired data
   b. Independent data
3. [if time] Introduction to power analysis

From last time

> Are swimmers faster in bodysuits than in regular swimsuits?

Below are the first few observations of the average velocity of competitive swimmers in a 1500m; one measurement was taken in a swimsuit, the other in a bodysuit.

swimmer	body.suit.velocity	swim.suit.velocity
1	1.57	1.49
2	1.47	1.37
3	1.42	1.35

- *two-sample* problem because there are two sets of observations
- observations are **paired** by swimmer

This is the easiest kind of two-sample problem because we can reduce it to a one-sample problem by performing inference on paired differences.

Inference for paired data

1. Calculate paired differences.

swimmer	body.suit.velocity	swim.suit.velocity	velocity.diff
1	1.57	1.49	0.08
2	1.47	1.37	0.1
3	1.42	1.35	0.07

2. Perform test as before.

```r
# test for paired difference
t.test(diffs, mu = 0, alternative = 'greater')
```

```
	One Sample t-test

data:  diffs
t = 12.318, df = 11, p-value = 4.443e-08
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 0.06620114        Inf
sample estimates:
mean of x
   0.0775
```

3. Report the result of the test. *You try.*

Formulating a two-sample problem

Two-sample problems are characterized by:

- one variable of interest
- two groups of observations
- objective to compare group means

Inference concerns the *difference in means*

$$\delta = \mu_1 - \mu_2$$

We just tested:

$$H_0 : \mu_{\text{body}} \leq \mu_{\text{swim}}$$

$$H_A : \mu_{\text{body}} > \mu_{\text{swim}}$$

Rearranging the data to emphasize two-sample problem structure:

swimmer	**suit**	**velocity**
1	body	1.57
1	swim	1.49
2	body	1.47
2	swim	1.37
3	body	1.42

- variable of interest: `velocity`
- grouping: `suit`
- pairing: `swimmer`

Hypotheses for two-sample tests

We can articulate two-sided and directional tests for the difference in means $\delta = \mu_1 - \mu_2$ and the corresponding interpretation in terms of the group means.

Difference in means

$$\text{two-sided}\begin{cases} H_0 : \delta \qquad 0 \\ H_A : \delta \qquad 0 \end{cases}$$

$$\text{lower-sided}\begin{cases} H_0 : \delta \qquad 0 \\ H_A : \delta \qquad 0 \end{cases}$$

$$\text{upper-sided}\begin{cases} H_0 : \delta \qquad 0 \\ H_A : \delta \qquad 0 \end{cases}$$

Group interpretation

$$\begin{cases} H_0 : \mu_1 \qquad \mu_2 \\ H_A : \mu_1 \qquad \mu_2 \end{cases}$$

$$\begin{cases} H_0 : \mu_1 \qquad \mu_2 \\ H_A : \mu_1 \qquad \mu_2 \end{cases}$$

$$\begin{cases} H_0 : \mu_1 \qquad \mu_2 \\ H_A : \mu_1 \qquad \mu_2 \end{cases}$$

Your turn: FAMuSS

> Does resistance training lead to greater strength gains on the nondominant arm?

ndrm.ch	drm.ch	sex	age	race	height	weight	actn3.r577x	bmi
40	40	Female	27	Caucasian	65	199	CC	33.11
25	0	Male	36	Caucasian	71.7	189	CT	25.84
40	0	Female	24	Caucasian	65	134	CT	22.3

Articulate and test an appropriate hypothesis for $\delta = \mu_{\text{ndrm}} - \mu_{\text{drm}}$

- Hypotheses:

  H_0 :

  H_A :

- Result:

Evolution of Darwin's finches

> Grant, P. (1986). Ecology and Evolution of Darwin's Finches, Princeton University Press, Princeton, N.J.

Peter and Rosemary Grant caught and measured all the birds from more than 20 generations of finches on the Galapagos island of Daphne Major.

- severe drought in 1977 limited food to large tough seeds
- selection pressure favoring larger and stronger beaks
- hypothesis: beak depth increased in 1978 relative to 1976

How do we test for a difference in the absence of pairing?

Finch beak data:

Year	Depth
1976	7.8
1976	9.5
1976	9.9
1978	10.3
1978	9.2
1978	10.9

- Variable: `Depth`
- Grouping: `Year`
- *No pairing!*

Inference for independent data

> Beak depths exemplify **independent data**: the groups of observations are unrelated.

Inference is based on the difference in group means:

$$T = \frac{\bar{x} - \bar{y}}{SE(\bar{x} - \bar{y})}$$

- $\bar{x}, \bar{y}$ are groupwise sample means
- $SE(\bar{x} - \bar{y}) = \sqrt{\frac{s_x^2}{n_x} + \frac{s_x^2}{n_y}}$
- degrees of freedom for t model are approximated

$$H_0 : \mu_{1976} \geq \mu_{1978}$$

$$H_A : \mu_{1976} < \mu_{1978}$$

```r
t.test(Depth ~ Year, data = finch,
       mu = 0, alternative = 'less')
```

```

	Welch Two Sample t-test

data:  Depth by Year
t = -4.5833, df = 172.98, p-value = 4.37e-06
alternative hypothesis: true difference in means
between group 1976 and group 1978 is less than 0
95 percent confidence interval:
      -Inf -0.427321
sample estimates:
mean in group 1976 mean in group 1978 
          9.469663          10.138202 
```

Two input formats

The **formula** format takes inputs:

1. an R formula
2. a data frame

```
# two-sample test (formula inputs)
t.test(Depth ~ Year, data = finch,
       mu = 0, alternative = 'less')
```

```
	Welch Two Sample t-test

data:  Depth by Year
t = -4.5833, df = 172.98, p-value = 4.37e-06
alternative hypothesis: true difference in means
between group 1976 and group 1978 is less than 0
95 percent confidence interval:
      -Inf -0.427321
sample estimates:
mean in group 1976 mean in group 1978
          9.469663          10.138202
```

`Depth ~ Year`: “depth *depends on* year”

The **vector** format takes inputs:

1. vector of observations for one group
2. vector of observations for the other group

```
# two-sample test (vector inputs)
t.test(depth76, depth78,
       mu = 0, alternative = 'less')
```

```
	Welch Two Sample t-test

data:  depth76 and depth78
t = -4.5833, df = 172.98, p-value = 4.37e-06
alternative hypothesis: true difference in means is
less than 0
95 percent confidence interval:
      -Inf -0.427321
sample estimates:
mean of x mean of y
 9.469663 10.138202
```

Interpreting results

```
    Welch Two Sample t-test

data:  Depth by Year
t = -4.5833, df = 172.98, p-value = 4.37e-06
alternative hypothesis: true difference in means
between group 1976 and group 1978 is less than 0
95 percent confidence interval:
      -Inf -0.427321
sample estimates:
mean in group 1976 mean in group 1978
          9.469663          10.138202
```

To report the results:

1. State conclusion
2. Interpret test result in context
3. Report statistics (*T*, *df*, *p*-value)
4. Provide point estimates

Careful about signs and directions!

Our findings suggest finch beak depth on Daphne Major increased as a result of natural selection following drought. The data provide strong evidence against the null hypothesis that mean beak depth remained comparable or diminished in the generation following the drought in favor of the alternative that mean beak depth increased (T = -4.58 on 172.98 degrees of freedom, p = 0.00000437). With 95% confidence, the increase in beak depth is estimated to be at least 0.427 mm, with a point estimate of 0.669 mm (SE = 0.1459).

Cloud data: paired or independent?

> Does dropping silver iodide onto clouds increase rainfall?

Data are rainfall measurements in a target area from 26 days when clouds were seeded and 26 days when clouds were not seeded.

- `rainfall` gives volume of rainfall in acre-feet
- `treatment` indicates whether clouds were seeded

Hypotheses to test:

$$H_0 : \mu_{\text{seeded}} \quad \mu_{\text{unseeded}}$$

$$H_A : \mu_{\text{seeded}} \quad \mu_{\text{unseeded}}$$

rainfall	treatment
978	Seeded
92.4	Seeded
242.5	Seeded
198.6	Seeded
830.1	Unseeded
4.9	Unseeded
87	Unseeded
81.2	Unseeded

Sleep drugs: paired or independent?

> Which (if either) of two soporific drugs is more effective?

Data are extra hours of sleep for 10 study participants when taking each of two drugs.

- `extra.sleep` gives hours of additional sleep relative to control
- `drug` indicates which sleep drug was taken
- `subject` indicates study participant id

Hypotheses to test:

$$H_0 : \mu_1 \quad \mu_2$$

$$H_A : \mu_1 \quad \mu_2$$

extra.sleep	drug	subject
0.7	1	1
1.9	2	1
-1.6	1	2
0.8	2	2
-0.2	1	3
1.1	2	3

Test assumptions

Inference relies on a t model providing a good approximation to the sampling distribution. This requires three assumptions:

1. variable of interest is numeric and not too discrete
2. observations are independent (besides pairing)
3. either:
    a. sample sizes are not too small
    b. or distribution(s) are symmetric and unimodal

Common issues:

Issue	**Consequence**
Highly discrete data	t model not appropriate
Dependent observations	SE is a biased estimate: nominal error rates and coverage are inaccurate
Small samples with *heavy* skew or *extreme* outliers	SE too small: inflated type I error and under-coverage

> *In each of these scenarios, different inference procedures should be used.*

Power calculations

> How much data do you need to collect in order to detect a difference of δ?

The statistical **power** of a test captures how often it detects a specified alternative.

- defined as $\beta = (1 - \text{type II error rate})$
- measures how often the test correctly rejects
- value depends on…
  a. magnitude of difference between null value and true value of parameter
  b. significance level
  c. sample size

```r
power.t.test(power = 0.95,
             delta = 0.5,
             sig.level = 0.05,
             type = 'two.sample',
             alternative = 'two.sided')
```

```
     Two-sample t test power calculation 

              n = 104.928
          delta = 0.5
             sd = 1
      sig.level = 0.05
          power = 0.95
    alternative = two.sided

NOTE: n is number in *each* group
```

$\Rightarrow$ need 105 observations in each group to detect a difference of 0.5 standard deviations at level 0.05 with type II error rate 5% or less

The equal-variance t-test

If it is reasonable to assume the (population) standard deviations are the same in each group, one can gain a bit of power (lower type II error rate) by using a different standard error:

$$SE_{\text{pooled}}(\bar{x} - \bar{y}) = \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}} \quad \text{where} \quad s_p = \underbrace{\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}}_{\text{weighted average of } s_x^2 \ \& \ s_y^2}$$

Implement by adding `var.equal = T` as an argument to `t.test()`.

- larger df is used, hence more frequent rejections
- avoid unless you have a small sample