

Stat 411/511

TWO-SAMPLE T-TEST

Jan 27 2012

Charlotte Wickham

stat511.cwick.co.nz

Homework 1

Solutions online.

Always put your name, and lab TA on your assignment.

Write full sentence answers, report units, use appropriate rounding.

(next time you'll lose marks for not doing these)

Review

One sample t -test

We compared two groups by looking at the **difference between pairs**

(a one-sample t -test on differences is called a paired t -test)

To identify if you should do a paired test ask yourself:

If I pick a subject in one group, does it have a natural partner in the other group?

(E.g. twins, left & right, before & after)

Big Idea from Wednesday

Facts about the **sampling distribution** allow us to use the **sample** to make **inferences** about the **population**.

We did the **one-sample** case, today we do the **two-sample** case.

Monday

Wednesday

Mechanism of chance	Random group allocation	Random sample from population (sometimes imaginary)
Distribution of the test statistic	Randomization distribution	Sampling distribution
The null distribution of the test statistic	Approximated with simulated data sets	Approximated by a mathematical curve

Today too

Two samples: Bumpus data

Display 2.1

p. 29

Humerus lengths (inches) of adult male house sparrows, 24 that perished and 35 that survived in a winter storm

<i>PERISHED</i>			<i>SURVIVED</i>
Average: .7279	9	65	Average: .7380
SD: .0235		66	SD: .0198
n: 24	9	67	n: 35
		68	
	932	69	
	3	70	
		71	
	96600	72	
		73	
	988761	74	
	543	75	
	422	76	
	5	77	
		78	

Legend: | 68 | 7 represents 0.687 inch

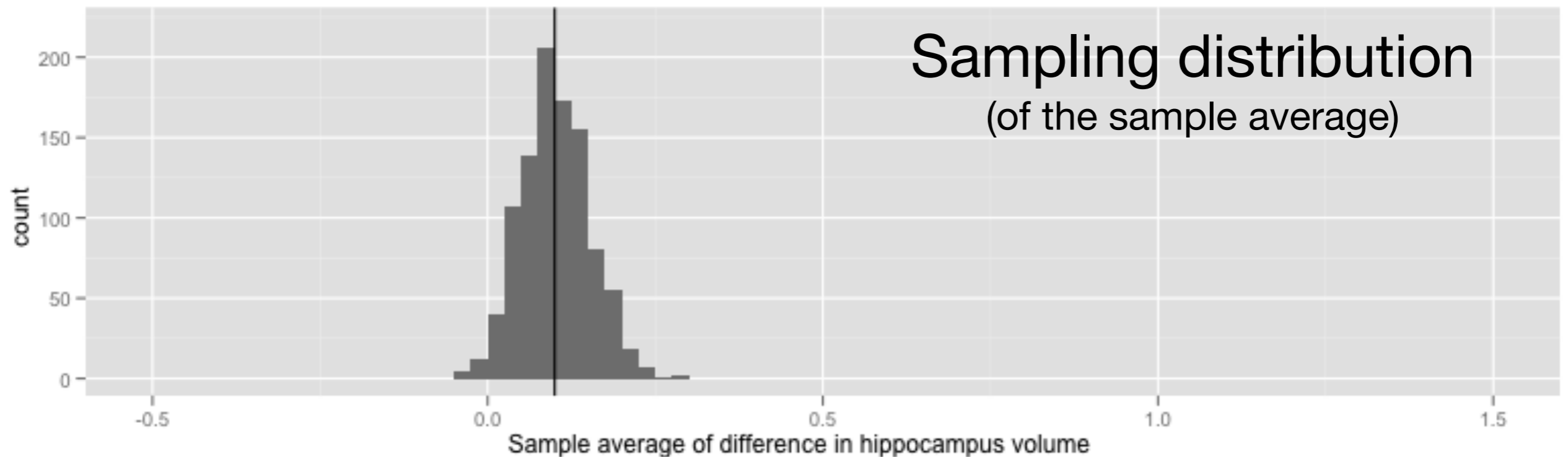
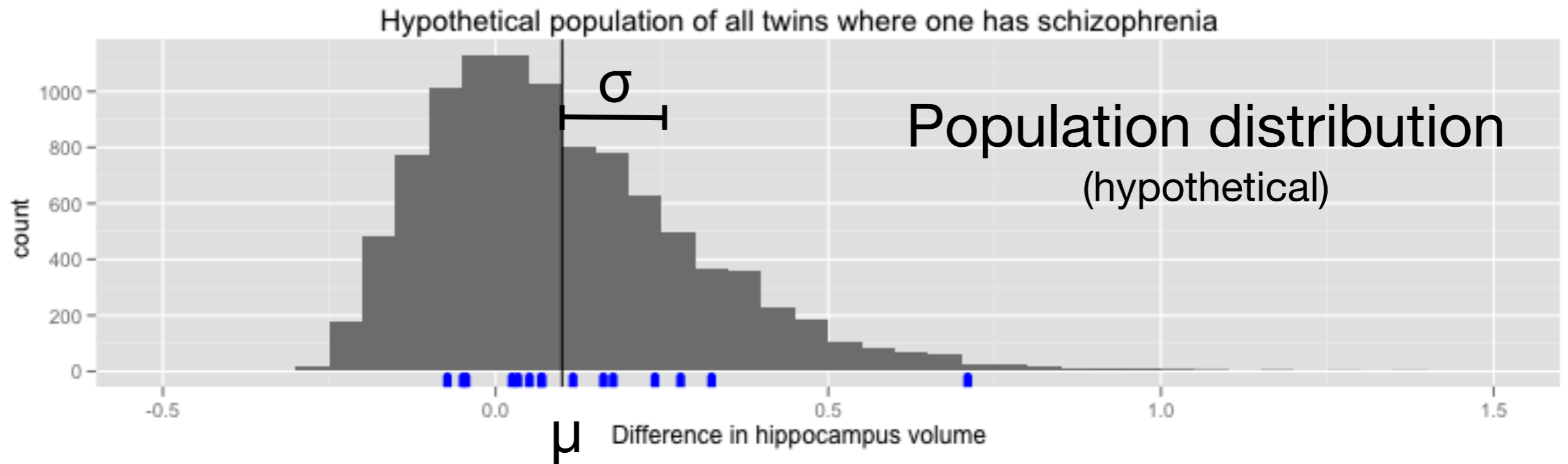
Is the population mean humerus length of the sparrows that survived different to the population mean of the sparrows that perished?

Your turn

In a one-sample test, we looked at the sampling distribution of the sample average.

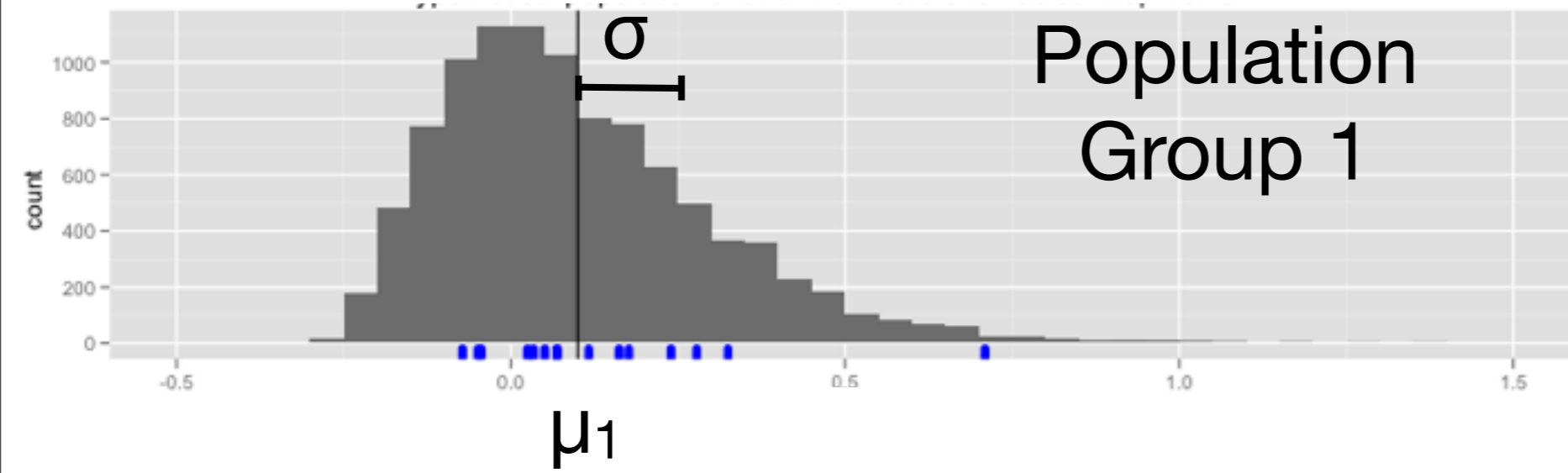
What would be a good one number summary for the two-sample problem?

From Wednesday one-sample case

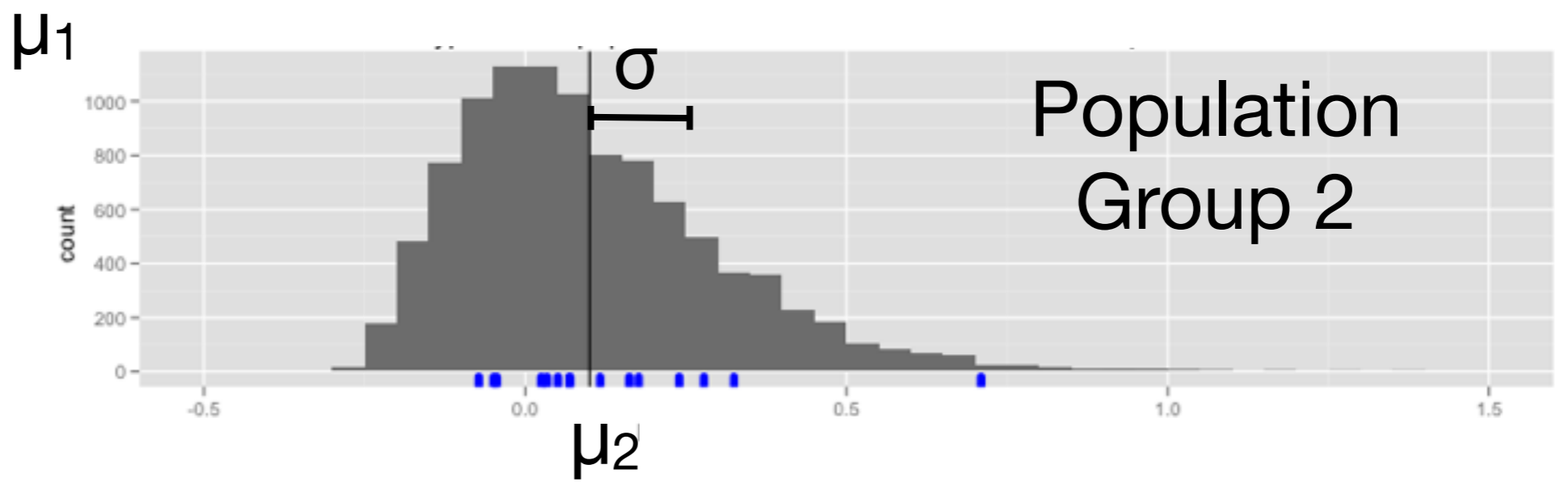
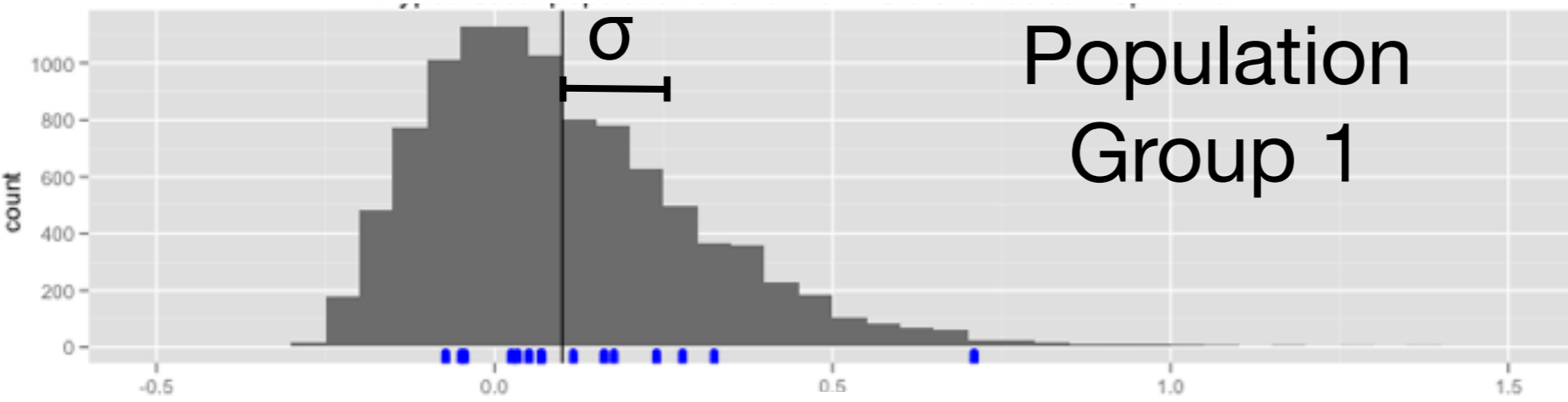


Two sample case

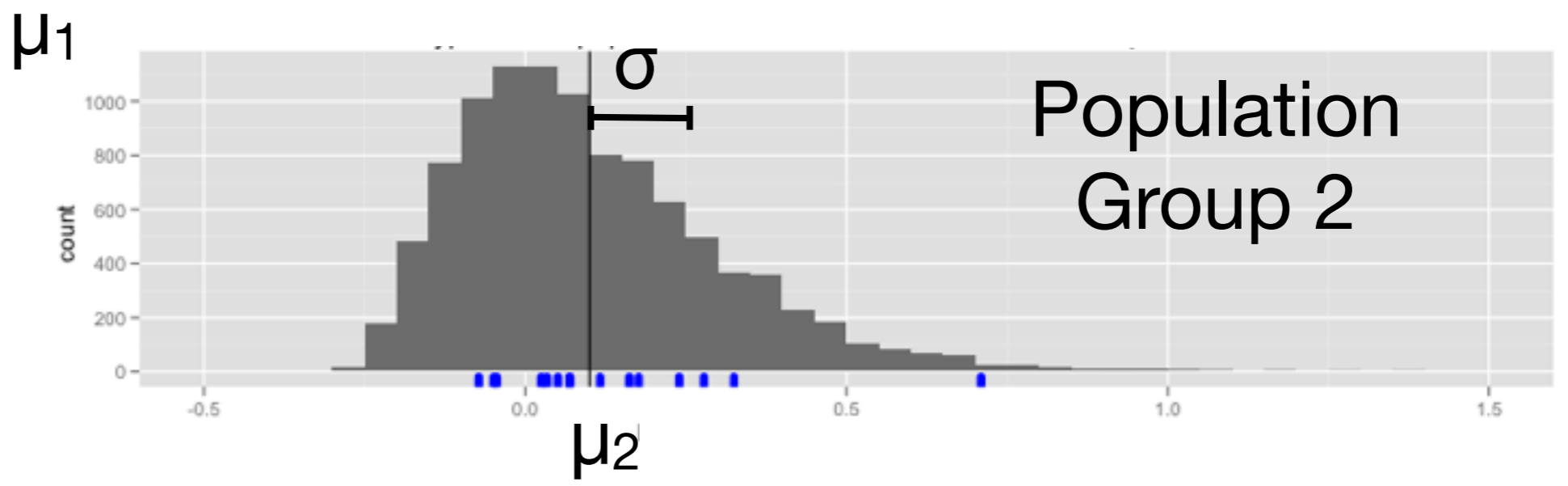
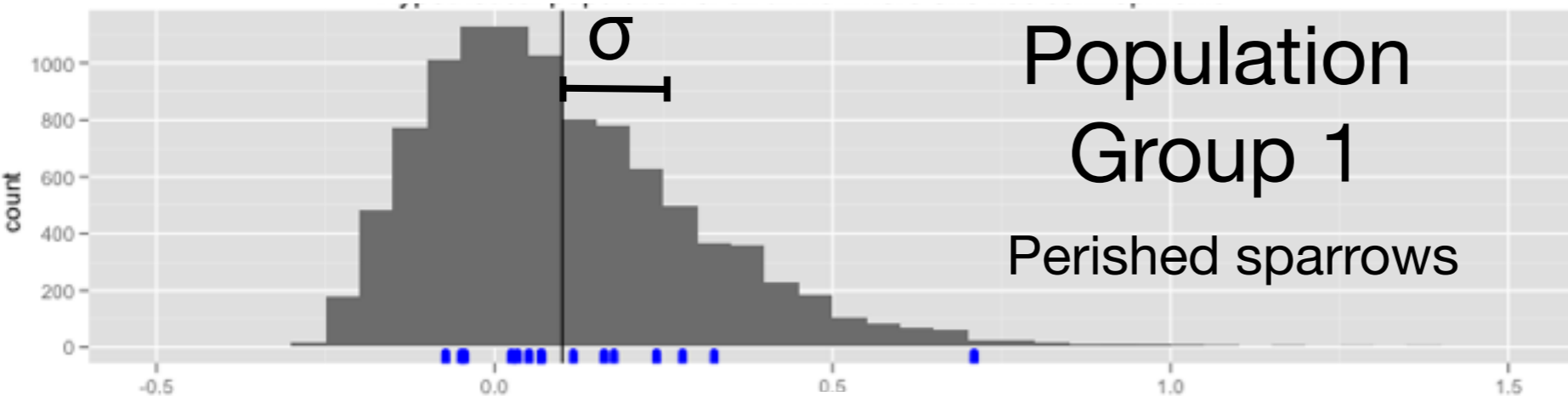
Two sample case



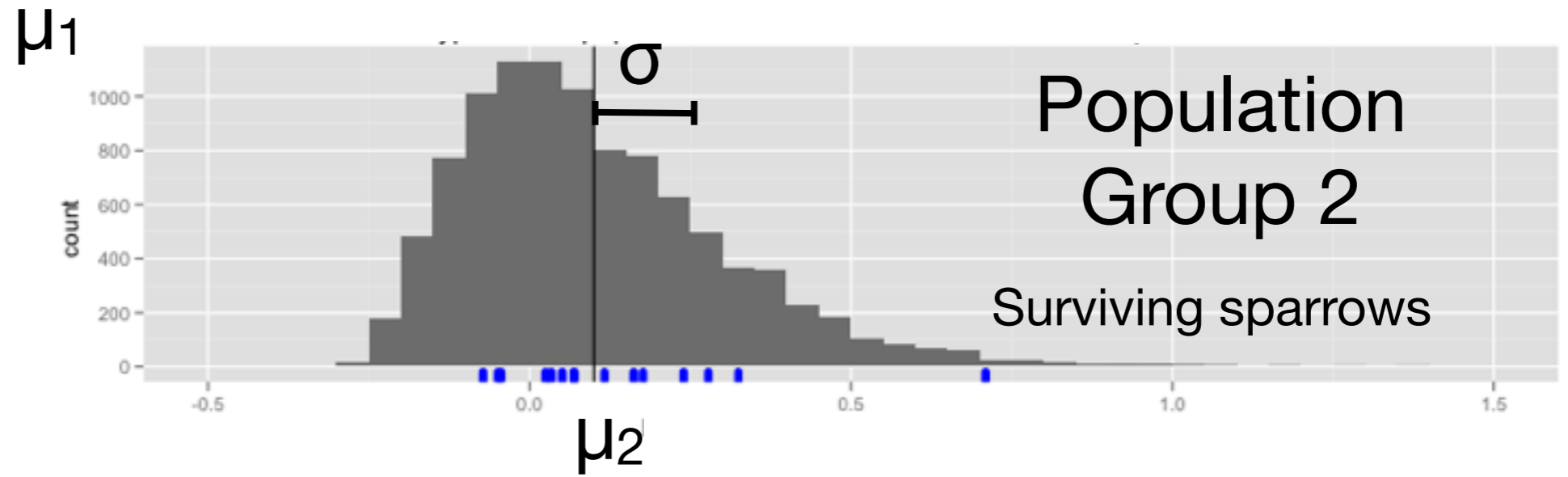
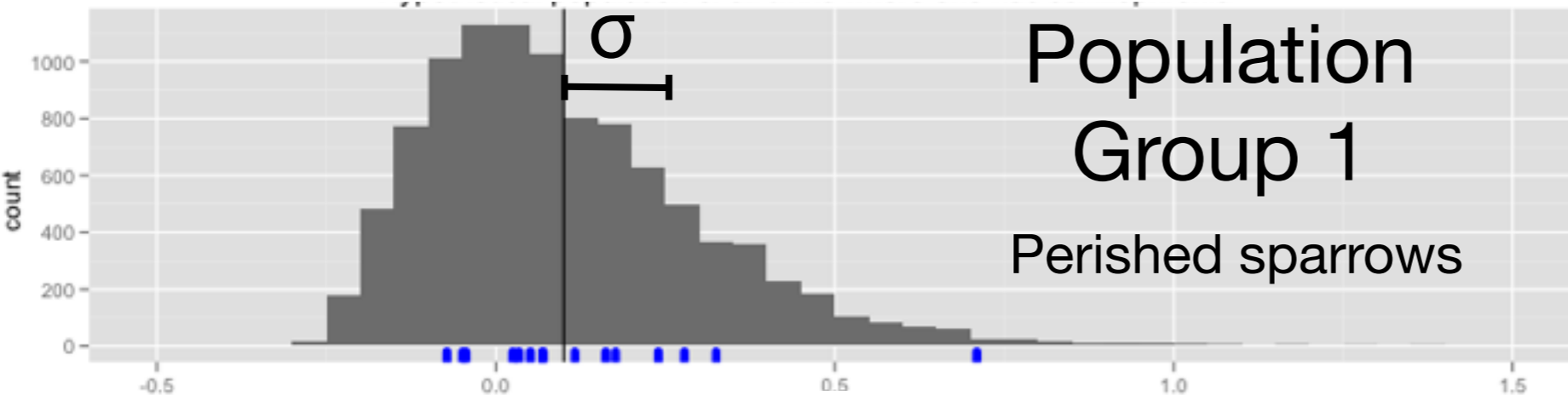
Two sample case



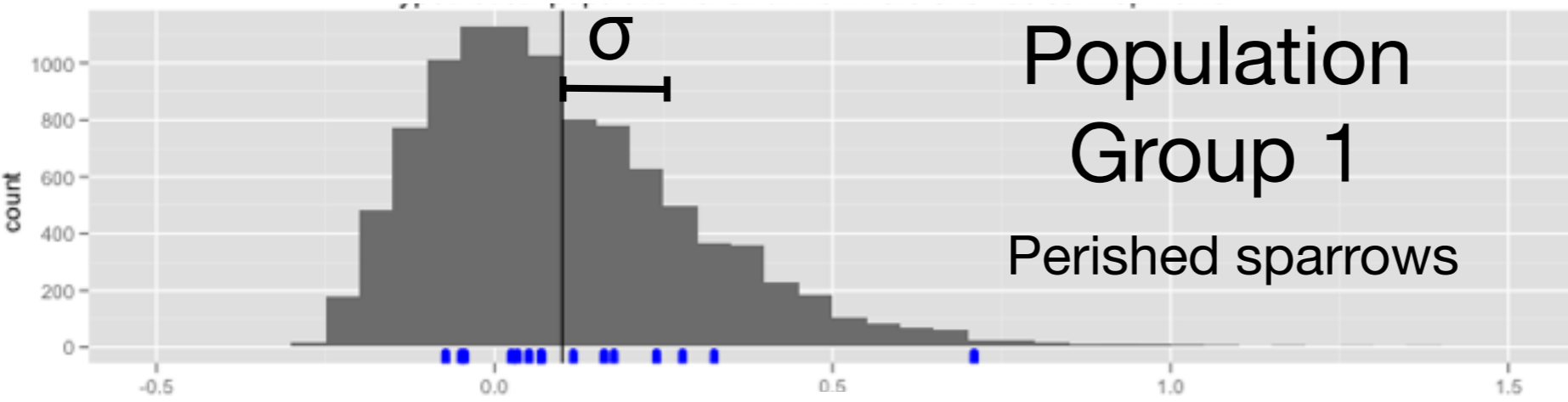
Two sample case



Two sample case



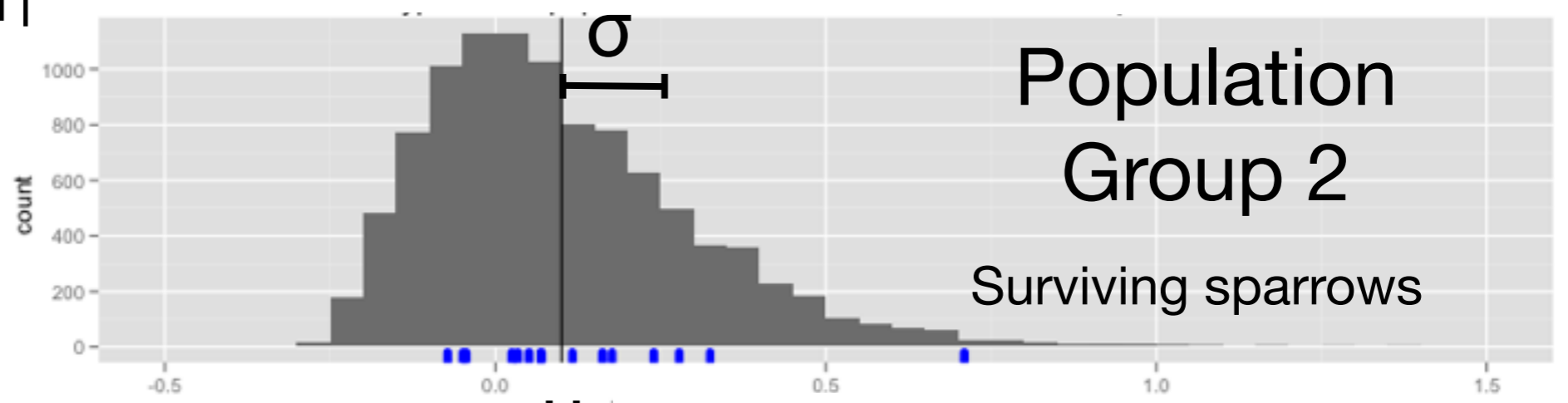
Two sample case



Take a sample of n_1 subjects and calculate average

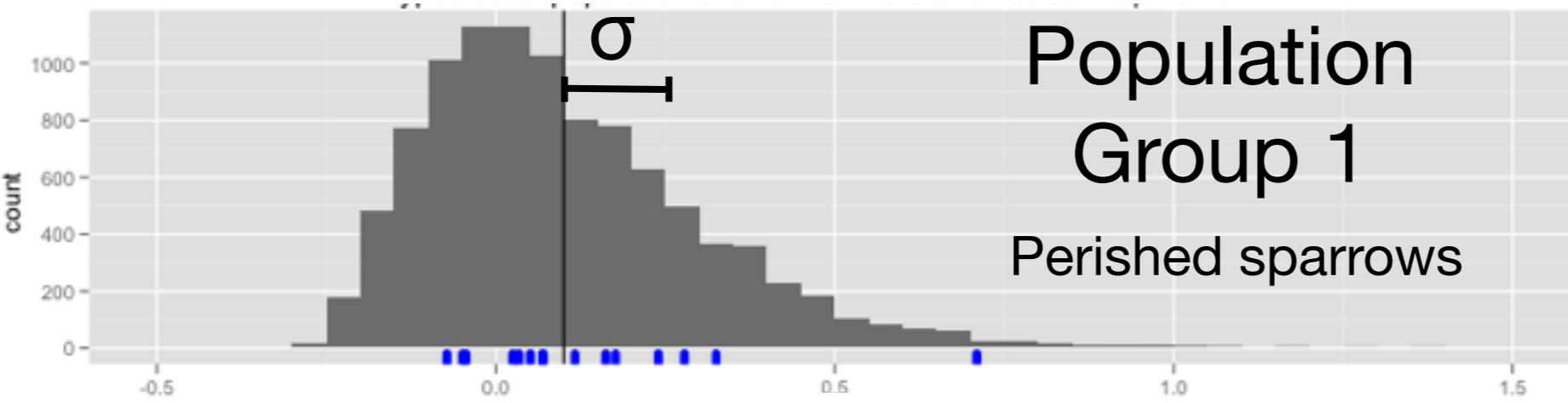
$$\bar{Y}_1$$

μ_1



μ_2

Two sample case



Population
Group 1
Perished sparrows

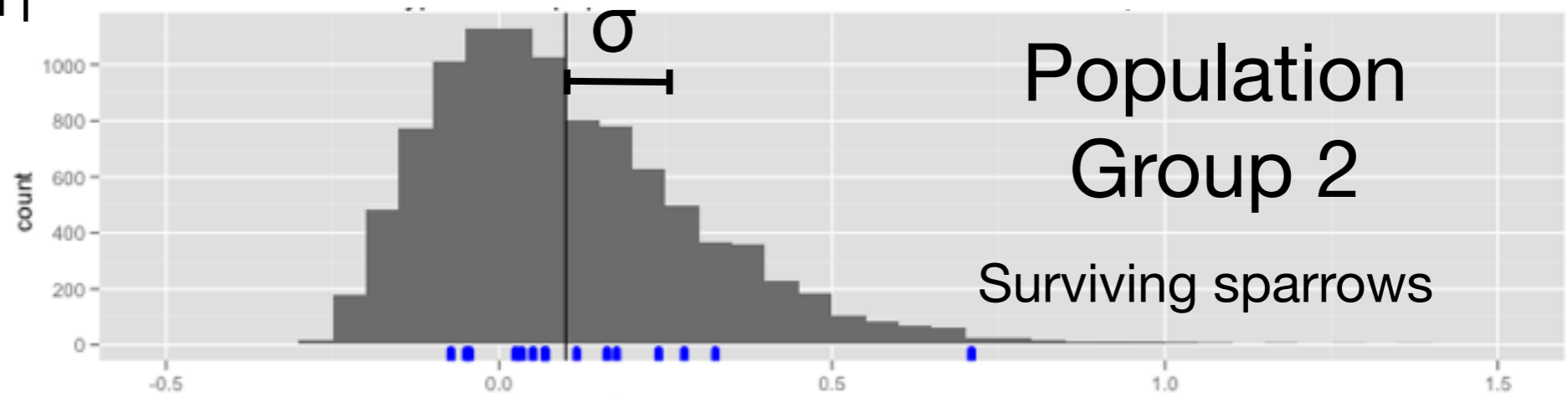
Take a sample of
 n_1 subjects and
calculate average

$$\overline{Y}_1$$

μ_1

Take a sample of
 n_2 subjects and
calculate average

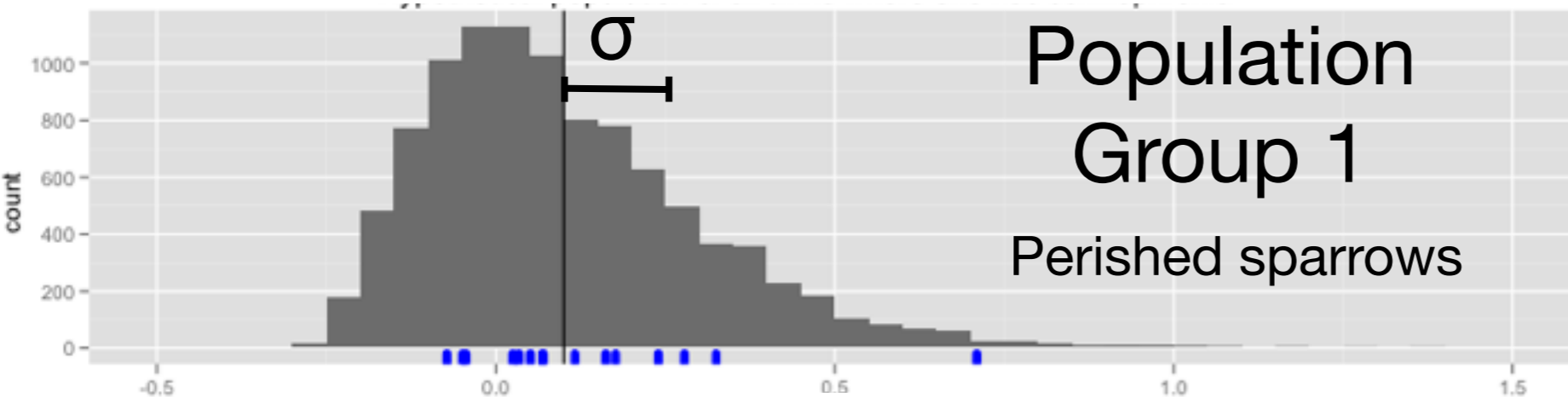
$$\overline{Y}_2$$



Population
Group 2
Surviving sparrows

μ_2

Two sample case



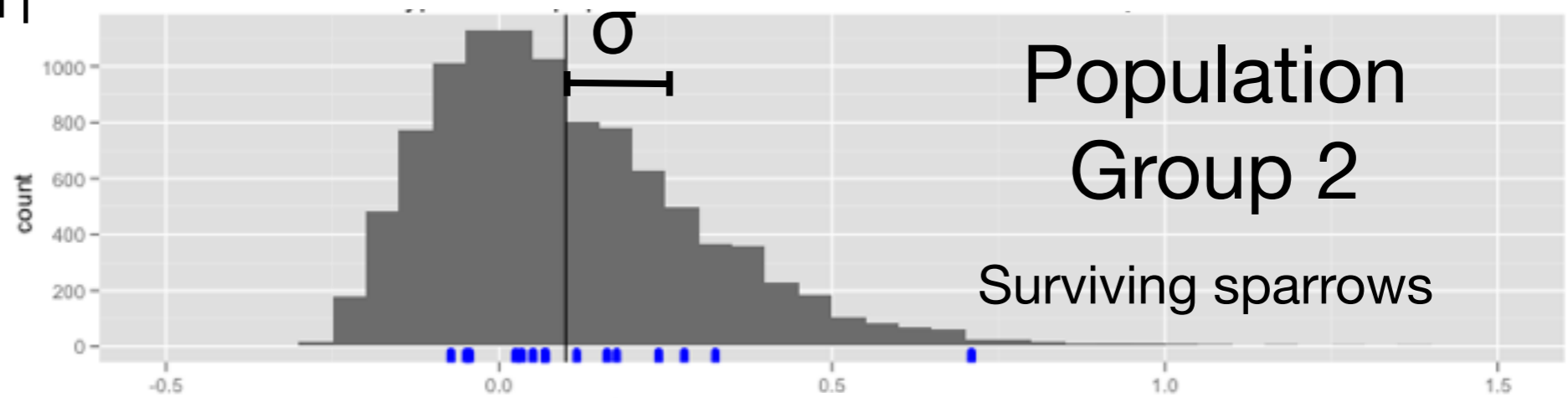
Take a sample of n_1 subjects and calculate average

$$\overline{Y}_1$$

μ_1

Take a sample of n_2 subjects and calculate average

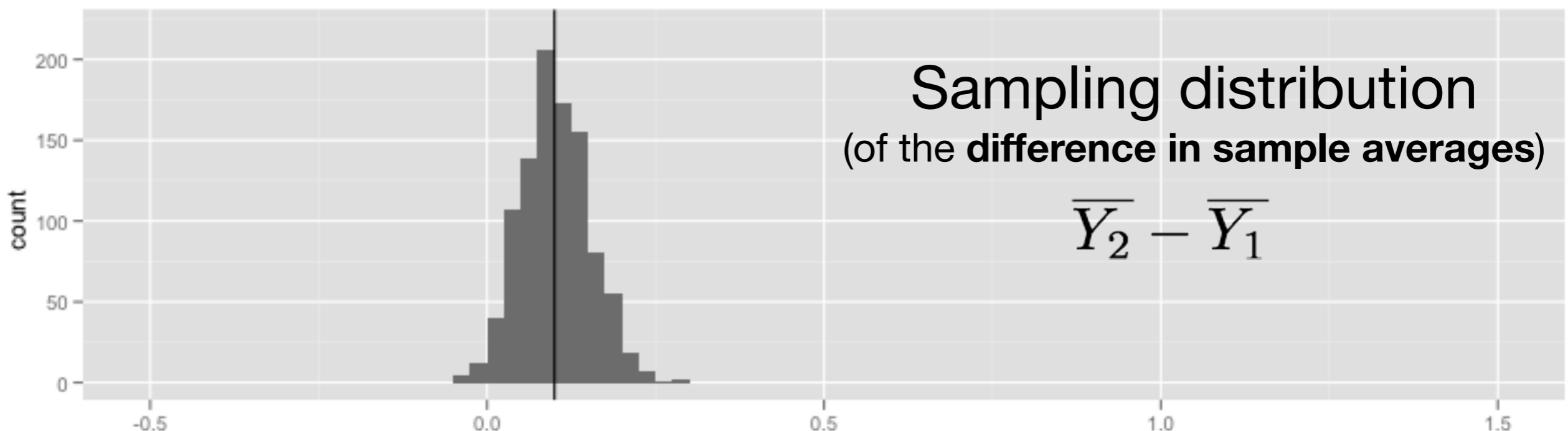
$$\overline{Y}_2$$



μ_2

Sampling distribution
(of the **difference in sample averages**)

$$\overline{Y}_2 - \overline{Y}_1$$



Use $\bar{Y}_2 - \bar{Y}_1$ to make inferences about
 $\mu_2 - \mu_1$

Assume that both groups have population distributions with the **same shape and standard deviation.**

Need to know the sampling distribution of $\bar{Y}_2 - \bar{Y}_1$

Very similar to the one sample case

Fact #4:

The two-sample t -ratio:

$$\frac{(\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1)}{SE_{\bar{Y}_2 - \bar{Y}_1}}$$

can be described by a **Student's t -distribution** with **$n_1 + n_2 - 2$** degrees of freedom

What is $SE_{\bar{Y}_2 - \bar{Y}_1}$?

An estimate of the standard deviation of the sampling distribution of $\bar{Y}_2 - \bar{Y}_1$

The standard deviation of $\bar{Y}_2 - \bar{Y}_1$ is

$$\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

We need to estimate σ

Pooled standard deviation

We have two samples each with their own standard deviation, s_1 and s_2 .

We need to combine them to get a pooled sample standard deviation.

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

Use to estimate σ

What is $SE_{\bar{Y}_2 - \bar{Y}_1}$?

$$s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Your turn

Find the standard error on the difference in averages for the Bumpus data

	Perished	Survived	
Mean	727.9	738.0	100ths of an inch
SD	23.5	19.8	
n	24	35	

$$SE_{\bar{Y}_2 - \bar{Y}_1} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

Two sample t -test

Null: the population means are the same, $\mu_1 = \mu_2$.

(or equivalently $\mu_1 - \mu_2 = 0$)

Alternative: the population means are different $\mu_1 \neq \mu_2$.

(or equivalently $\mu_1 - \mu_2 \neq 0$)

(or a one sided alternative)

Steps in a two-sample t -test

1. Find the averages and standard deviations ✓
2. Figure out the degrees of freedom
($n_1 + n_2 - 2$)
3. Find the standard error on the difference in averages ✓
4. Calculate the t -ratio under the null ($\mu_1 - \mu_2 = 0$)

$$\frac{(\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1)}{SE_{\bar{Y}_2 - \bar{Y}_1}}$$

6. Compare it to a Student's t -distribution with d.f degrees of freedom

$$2 * (1 - pt(t.statistic, d.f))$$

Your turn

Complete the two-sample t -test for the Bumpus data.

```
t.test(peri, surv, var.equal = TRUE)
```

Two Sample t-test

```
data:  peri and surv
```

```
t = -1.777, df = 57, p-value = 0.0809
```

```
alternative hypothesis: true difference in means  
is not equal to 0
```

```
95 percent confidence interval:
```

```
-21.446053    1.279386
```

```
sample estimates:
```

```
mean of x mean of y
```

```
727.9167  738.0000
```

```
t.test(peri, surv, var.equal = TRUE)
```

Two Sample t-test

```
data: peri and surv
```

```
t = -1.777, df = 57, p-value = 0.0809
```

```
alternative hypothesis: true difference in means  
is not equal to 0
```

```
95 percent confidence interval:
```

```
-21.446053    1.279386
```

```
sample estimates:
```

```
mean of x mean of y
```

```
727.9167    738.0000
```

next Monday