

# Inference for Means II

## IQ Example, two-sample

Consider a data set containing the IQs of 36 men and the IQs of 36 women. Let's use this data to test the following.

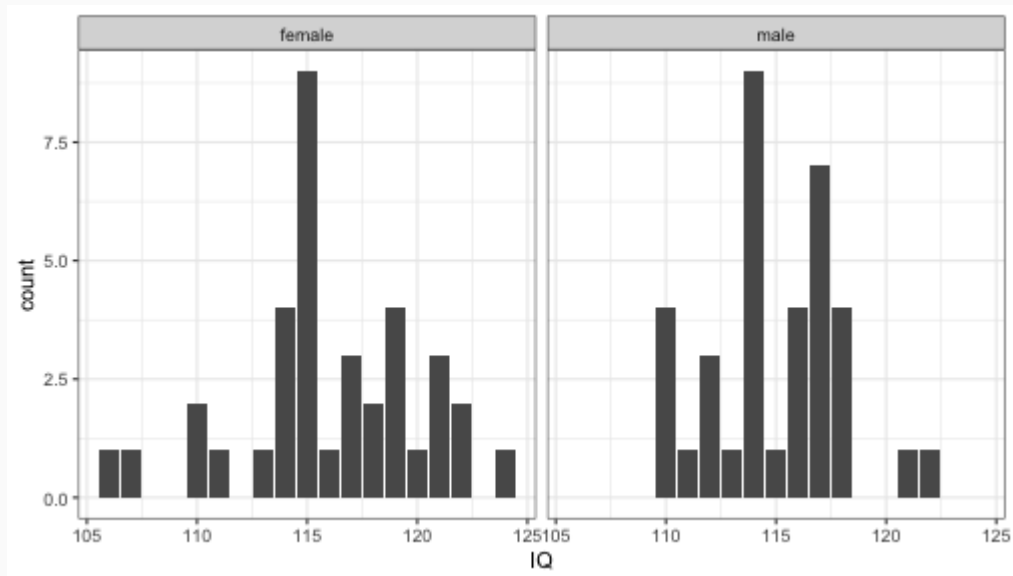
$$H_0 : \mu_M - \mu_F = 0$$

$$H_A : \mu_M - \mu_F \neq 0$$

##		IQ	sex
##	72	122	female
##	38	119	female
##	46	115	female
##	7	116	male
##	14	117	male
##	21	114	male

```
(ds <- d %>%
  group_by(sex) %>%
  summarize(mean = mean(IQ),
            s = sd(IQ),
            n = n()))
```

```
## # A tibble: 2 x 4
##   sex      mean      s      n
##   <fct>  <dbl> <dbl> <int>
## 1 female  116.   4.09   36
## 2 male    115.   2.97   36
```



# Two sample t-test, step 1

## Calculate observed test statistic

```
x_bar_f <- ds[1, 2] %>% pull()
x_bar_m <- ds[2, 2] %>% pull()
s_f <- ds[1, 3] %>% pull()
s_m <- ds[2, 3] %>% pull()
n_f <- ds[1, 4] %>% pull()
n_m <- ds[2, 4] %>% pull()
```

```
t_obs <- ((x_bar_m - x_bar_f) - 0) /
  sqrt(s_m^2/n_m + s_f^2/n_f)
t_obs
```

```
## [1] -1.29
```

# Two sample t-test, step 2

## Check conditions

- Nearly normal populations (barplots looked OK)
- Independent observations
  - Within each group
  - Between group means

# Two sample t-test, step 3

## Compute p-value

```
df <- min(n_f - 1, n_m - 1)
pt(t_obs, df = df)
```

```
## [1] 0.104
```

```
pt(t_obs, df = df) * 2
```

```
## [1] 0.207
```

Our data is consistent with the notion that men and women have the same mean IQ.

# Two-sample "t-test" via permutation

## Idea

A slightly different way to lay out the null hypothesis is to say that both men's and women's IQs were drawn from the same distribution and thus the two variables, `IQ` and `sex`, are independent of one another.

With a null of "`independence`", we can generate data using *permutation*.

## Two-sample "t-test" via permutation, cont.

```
d %>%  
  specify(IQ ~ sex) %>%  
  hypothesize(null = "independence") %>%  
  generate(reps = 500, type = "permute") %>%  
  calculate(stat = "diff in means",  
            order = c("male", "female"))
```

```
## # A tibble: 500 x 2  
##   replicate    stat  
##   <int>    <dbl>  
## 1         1  0.194  
## 2         2  0.417  
## 3         3 -0.139  
## 4         4  0.194  
## 5         5 -0.583  
## 6         6 -0.361  
## 7         7 -1.14  
## 8         8  0.472  
## 9         9  0.361
```



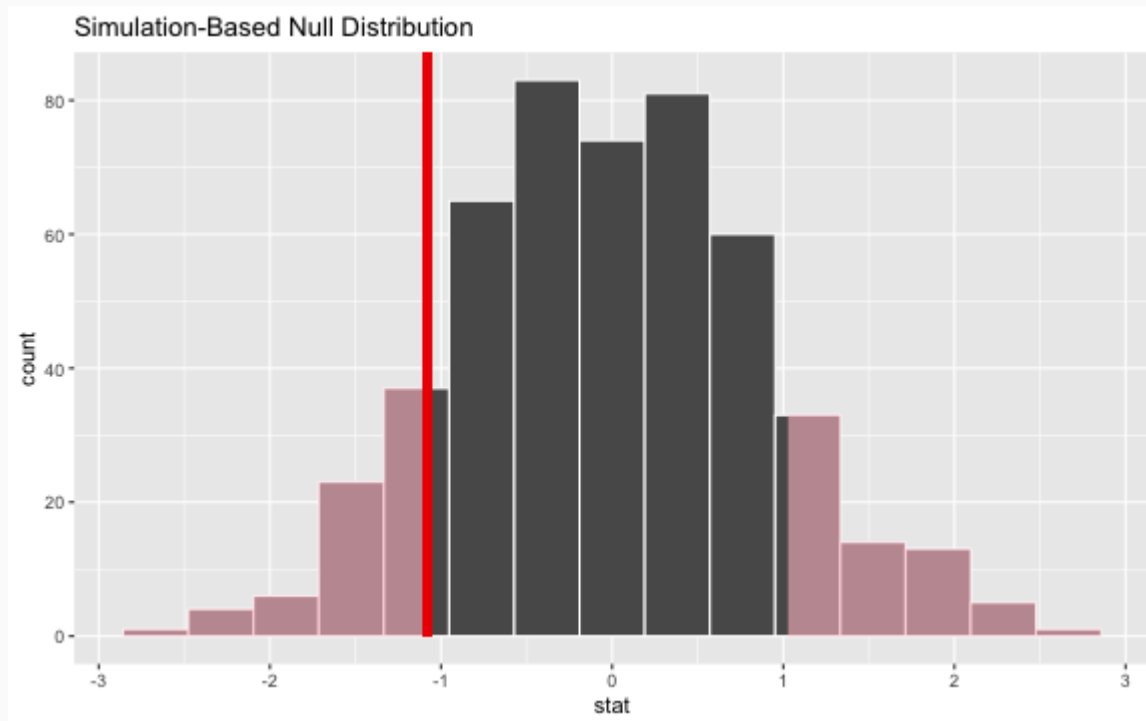
```
null <- d %>%  
  specify(IQ ~ sex) %>%  
  hypothesize(null = "independence") %>%  
  generate(reps = 500, type = "permute") %>%  
  calculate(stat = "diff in means",  
            order = c("male", "female"))
```

```
obs_stat <- d %>%  
  specify(IQ ~ sex) %>%  
  # hypothesize(null = "independence") %>%  
  # generate(reps = 500, type = "permute") %>%  
  calculate(stat = "diff in means",  
            order = c("male", "female")) %>%  
  pull()  
obs_stat
```

```
## [1] -1.08
```

# Visualizing the null

```
null %>%  
  visualize() +  
  shade_p_value(obs_stat = obs_stat,  
                direction = "both")
```



## Computing a p-value

```
null %>%  
  get_p_value(obs_stat = obs_stat,  
              direction = "both")
```

```
## # A tibble: 1 x 1  
##   p_value  
##   <dbl>  
## 1    0.232
```

Our data is consistent with the notion that men and women have the same mean IQ.

## Selecting a method

- Both methods lead to slightly different p-values but the same ultimate conclusion.
- Select based on which assumptions seem more reasonable.
  - The  $t$ -distribution assumes normality of IQs within each group.
  - The permutation method has a "stronger" null: that the whole distributions, not just the means, are the same.

But more important than both of these things: *be sure you understand the provenance of the data.*

# Original Data

Data were collected from schools in a large city on a set of thirty-six children who were identified as gifted children soon after they reached the age of four.

```
head(gifted)
```

##	score	fatheriq	motheriq	speak	count	read	edutv	carto
## 1	159	118	121	18	26	1.9	3.00	2
## 2	164	117	119	20	37	2.5	1.75	3
## 3	154	112	115	20	32	2.2	2.75	2
## 4	157	112	115	12	24	1.7	2.75	2
## 5	156	117	118	17	34	2.2	2.25	2
## 6	150	114	114	13	28	1.9	1.25	3

## Paired data

If there is a natural pairing between observations in two groups of size  $n$ , it can make more sense to analyze them as a *single* sample of  $n$  differences.

```
gifted %>%  
  mutate(diff = fatheriq - motheriq) %>%  
  select(fatheriq, motheriq, diff)
```

##	fatheriq	motheriq	diff
## 1	118	121	-3
## 2	117	119	-2
## 3	112	115	-3
## 4	112	115	-3
## 5	117	118	-1
## 6	114	114	0
## 7	116	115	1
## 8	111	113	-2
## 9	110	110	0

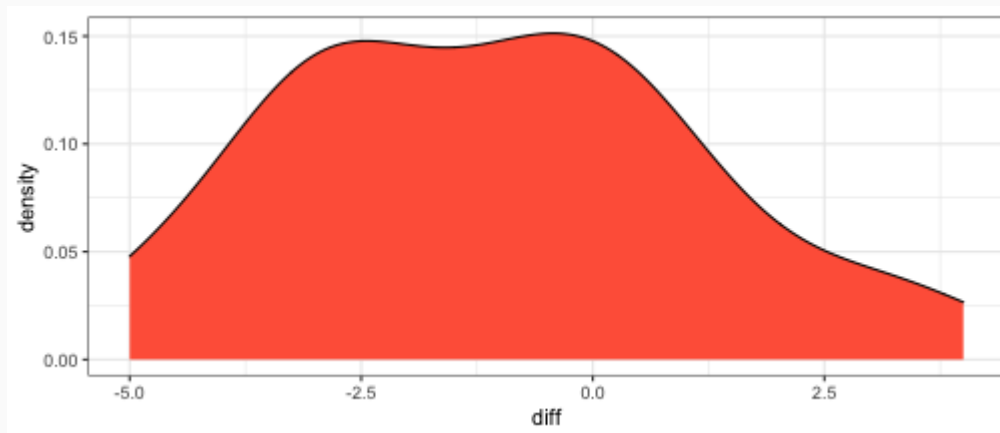
# Paired t-test

$$H_0 : \mu_{diff} = 0$$

$$H_A : \mu_{diff} \neq 0$$

## Check conditions

1. Independent observations
2. Nearly normal population



# Paired t-test (cont.)

## Compute a test statistic

```
(gs <- gifted %>%  
  mutate(diff = fatheriq - motheriq) %>%  
  summarize(mean = mean(diff), s = sd(diff), n = n()))
```

```
##      mean      s  n  
## 1 -1.08  2.23 36
```

```
(t_obs <- (gs$mean - 0)/sqrt(gs$s^2/gs$n))
```

```
## [1] -2.91
```

- $df = n - 1$



## Paired compared

```
sqrt(gs$s^2/gs$n)
```

```
## [1] 0.372
```

```
sqrt(s_m^2/n_m + s_f^2/n_f)
```

```
## [1] 0.843
```

While the mean difference is the same in the paired and independent tests, if the data is paired, the dependency leads to a smaller SE.

This principle is widely used in experiment design, e.g. pre- and post-test.