

# Statistics 305/605: Introduction to Biostatistical Methods for Health Sciences

## Chapter 11 : Inference for Two Means

Jinko Graham

2018-09-13

# Comparison of Two Means (Chapter 11)

## Context

- ▶ We have measurements sampled from two populations.
- ▶ Want to make inference about the difference between the populations.
- ▶ In particular, interested in the difference between the two population means, denoted  $\mu_1$  and  $\mu_2$ .

## Notation

- ▶ Let  $x_{11}, \dots, x_{1n_1}$  denote a sample from the first population and  $x_{21}, \dots, x_{2n_2}$  denote a sample from the second.
- ▶ The sample averages  $\bar{x}_1$  and  $\bar{x}_2$  estimate the population means  $\mu_1$  and  $\mu_2$ , respectively; so  $\bar{x}_1 - \bar{x}_2$  estimates  $\mu_1 - \mu_2$ .
- ▶ We're interested in confidence intervals for  $\mu_1 - \mu_2$  and tests of  $H_0 : \mu_1 - \mu_2 = 0$  vs.  $H_a : \mu_1 - \mu_2 \neq 0$ .

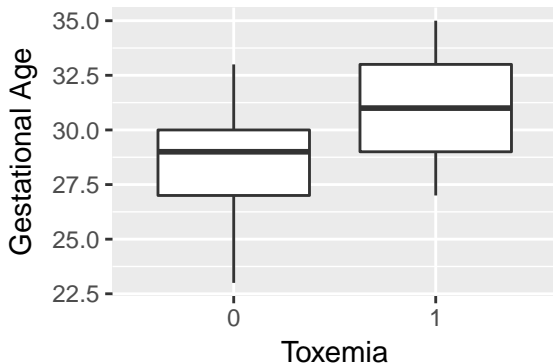
## Example: Low Birthweight Infants

- ▶ Data on 100 infants with birth weight less than 1500g.
  - ▶ Variables are: head circumference (cm), birth length (cm), gestational age (wks), birth weight (g), mother's age (yrs), and mother's status for toxemia (1=high blood pressure during pregnancy, 0=not)
- ▶ Compare the distribution of variables such as age and birth weight in moms with to mom's without toxemia
- ▶ The first few rows of the data set are as follows:

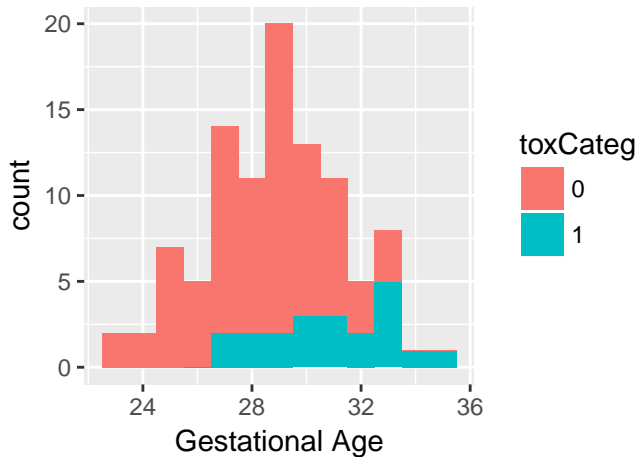
##	headcirc	length	gestage	birthwt	momage	toxemia
## 1	27	41	29	1360	37	0
## 2	29	40	31	1490	34	0
## 3	30	38	33	1490	32	0
## 4	28	38	31	1180	37	0
## 5	29	38	30	1200	29	1
## 6	23	32	25	680	19	0

## Gestational Age by Toxemia

- ▶ Question: Does the distribution of gestational age differ in moms with toxemia vs. moms without toxemia?
- ▶ Explore differences by toxemia status graphically, using boxplots (below) and histograms (next slide) in the sample.



## Gestational Age by Toxemia: Histograms



## Gestational Age by Toxemia: Summary Statistics

- ▶ The sample means and SDs of gestational age for each toxemia category are summarized below.

```
## # A tibble: 2 x 3
##   toxCateg mean    sd
##   <fct>    <dbl> <dbl>
## 1 0        28.4  2.32
## 2 1        30.9  2.32
```

- ▶ The sample means of the gestational ages differ between the toxemia groups, but the sample SDs look the same.
- ▶ Could the difference in the sample means be due to chance?

# Outline of Inference Approach

- ▶ Same basic approach to inference as in the one-sample problem:
  - ▶ Inference is based on the sampling distribution of the statistic  $\bar{X}_1 - \bar{X}_2$
- ▶ Transform  $\bar{X}_1 - \bar{X}_2$  to a *pivotal quantity*,  $Z$ , if population SDs  $\sigma_1$  and  $\sigma_2$  are known.
- ▶ When  $\sigma$ 's are unknown, as is typically the case, we substitute estimates to obtain a *pivotal quantity*  $T$ .
- ▶ Confidence intervals and hypothesis tests follow from the sampling distribution of  $T$ .
- ▶ Note: We omit the following topics in this course:
  - ▶ Paired-samples  $t$ -test (Section 11.1 of text)
  - ▶ Two-sample  $t$ -test assuming equal SDs (Section 11.2.1)

## Sampling Distribution of $\bar{X}_1 - \bar{X}_2$

- ▶ We have simple random samples (SRSs) of size  $n_1$  for group 1 and  $n_2$  for group 2.
- ▶ These samples are independent.
- ▶ Then the distribution of  $\bar{X}_1 - \bar{X}_2$  has
  - ▶ mean  $\mu_1 - \mu_2$  and
  - ▶ SD  $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ , where  $\sigma_1$  and  $\sigma_2$  are the population SDs for group 1 and group 2, respectively.
- ▶ If the sample sizes are large enough, the CLT tells us that the shape of this distribution is approximately normal.



## Z Transformation

- ▶ “Standardizing” a random variable by subtracting its population mean and dividing by its population SD gives a random variable with mean 0 & SD 1.
- ▶ For normal random variables, the transformation does not change the distribution; the standardized random variable is still normally distributed.
- ▶ Conclude that if  $\bar{X}_1 - \bar{X}_2$  is approximately normal with mean  $\mu_1 - \mu_2$  and SD  $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ , then

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1),$$

where  $\sim$  means “is distributed as”.

# T Transformation

- ▶ Inserting sample SDs  $s_1$  and  $s_2$  for the parameters  $\sigma_1$  and  $\sigma_2$  in  $Z$  gives

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- ▶ What is the distribution of  $T$ ?
  - ▶ Turns out we can approximate it by a  $t$ -distribution with  $\nu$  df.
  - ▶ We won't ever use the formula for  $\nu$  but if you're curious it is given on page 270 of the text.
- ▶ Instead, computer software such as R automatically calculates  $\nu$  for us, as shown next.

## Illustration with Gestational Age and Toxemia

- ▶ The following is the output of the `t.test()` function for these data (see the R demo for details):

```
##  
##  Welch Two Sample t-test  
##  
## data:  gestage by toxCateg  
## t = -4.4745, df = 31.465, p-value = 9.365e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
##  -3.712089 -1.388574  
## sample estimates:  
## mean in group 0 mean in group 1  
##           28.35443           30.90476
```

- ▶ The df is  $\nu = 31.465$ .
- ▶ We can also see that the software computed a 95% confidence interval and p-value. More on these in a few slides.

## Confidence Intervals for $\mu_1 - \mu_2$

- ▶ The level- $C$  CI for  $\mu_1 - \mu_2$  is of the form

estimate  $\pm$  margin of error

- ▶ The estimate is  $\bar{x}_1 - \bar{x}_2$
- ▶ The margin of error is  $t^* \times SE$  where
  - ▶  $t^*$  is the upper  $(1 - C)/2$  critical value of the  $t$ -distribution with  $\nu$  df, and
  - ▶  $SE$  is the **estimated SD** of  $\bar{x}_1 - \bar{x}_2$ :  $\sqrt{s_1^2/n_1 + s_2^2/n_2}$

## Application to Gestational Age and Toxemia

- ▶ Calculate a 90% CI for the difference between mean gestational age in the toxemia and non-toxemia groups.
- ▶ The relevant sample statistics to four digits are as follows:

group	sample mean ( $\bar{x}$ )	sample SD ( $s$ )	sample size ( $n$ )
1: non-toxemia	28.35	2.321	79
2: toxemia	30.90	2.322	21

- ▶ Estimate is  $\bar{x}_1 - \bar{x}_2 = 28.35 - 30.90 = -2.55$
- ▶ Margin of error is  $t^* \times SE$  where
  - ▶ computer software calculates a critical value of  $t^* = 1.695$  (see the R demo).
  - ▶ the SE is
$$\sqrt{s_1^2/n_1 + s_2^2/n_2} = \sqrt{2.321^2/79 + 2.322^2/21} = 0.570.$$
  - ▶ Hence the margin of error is  $1.696 \times 0.570 = 0.966$ .
- ▶ CI is  $(-2.55 - 0.966, -2.55 + 0.966) = (-3.516, -1.584)$ .

# Interpretation

- ▶ The 90% CI is approximately  $(-3.5, -1.6)$ .
- ▶ The text suggests an interpretation such as:
  - ▶ “90% of intervals constructed in this way cover the true difference between mean gestational age in the **non-toxemia** and **toxemia** groups.”
- ▶ Another common style of interpretation is:
  - ▶ “We are 90% confident that the true difference between mean gestational age in the **non-toxemia** and **toxemia** groups is between  $-3.5$  and  $-1.6$ .”
- ▶ Or, we might find it more natural to switch the order of the groups, which would switch the sign of the difference:
  - ▶ “We are 90% confident that the true difference between mean gestational age in the **toxemia** and **non-toxemia** groups is between  $1.6$  and  $3.5$ ”

# Hypothesis Test

- ▶ For the null hypothesis  $H_0 : \mu_1 - \mu_2 = 0$  of no difference between the groups, the formula for the observed  $t$ -statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- ▶ Observed values of  $t$  that are extreme in the sense of being more compatible with  $H_a$  are taken as evidence against  $H_0 : \mu_1 - \mu_2 = 0$ .
- ▶ The  $p$ -value is the chance of a value that is as or more extreme than what we observed, under  $H_0$ .
- ▶ Taking  $T$  to have a  $t$  distribution on  $\nu$  df, we get a  $p$ -value of  $p = 2P(T \geq |t|)$ . for  $H_a : \mu_1 - \mu_2 \neq 0$ .

## Application to Gestational Age and Toxemia

- ▶ We can re-use the calculations from the CI example (page 13), for the difference in sample means ( $\bar{x}_1 - \bar{x}_2 = -2.55$ ) and the SE of the difference ( $\sqrt{s_1^2/n_1 + s_2^2/n_2} = 0.570$ ):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{-2.55}{0.570} = -4.474.$$

- ▶ Computer software (R demo) returns a  $p$ -value of  $9.4 \times 10^{-5}$ , or 0.000094, for a 2-sided alternative hypothesis.



# Interpretation

- ▶ Small  $p$ -values (e.g.  $< .05$ ) are taken as evidence against the null hypothesis.
- ▶ Our  $p$ -value of  $9.4 \times 10^{-5}$  is very small.
- ▶ If we had set a level of  $\alpha = 10\%$  for the test, we'd declare that:  
*"We reject the null hypothesis that the mean gestational age is the same in the toxemia and non-toxemia groups at the 10% level."*
- ▶ If we hadn't set a level of the test in advance, we might report our results as:  
*"There is strong evidence that the mean gestational age is different in the toxemia and non-toxemia groups ( $p < 0.001$ )."*

# Cause and Effect

- ▶ Our two-sample t test has revealed that toxemia and gestational age are **associated**.
  - ▶ The distribution of gestational age is different in the two toxemia groups (different means, lower in toxemia group).
- ▶ But, an association does **not** mean that toxemia has a causal effect on gestational age.
  - ▶ It could be that gestational age affects toxemia.
  - ▶ Or, there could be a hidden *confounding variable* that affects both toxemia and gestational age, that accounts for their association.

(More on confounding later, when we study multiple regression.)

# Relationship Between Confidence Intervals and Tests

- ▶ In the low birth weight example, the 90% CI does not cover zero, and the hypothesis test of  $H_0 : \mu_1 - \mu_2 = 0$  vs  $H_a : \mu_1 - \mu_2 \neq 0$  rejects the null hypothesis at the 10% level.
- ▶ Conversely, when a 90% CI **does** cover zero, the corresponding test against a two-sided alternative will **retain** the null hypothesis at the 10% level.
- ▶ This is a general property of tests of a population parameter  $\theta$ .
  - ▶ A level- $\alpha$  test of  $H_0 : \theta = \theta_0$  versus  $H_a : \theta \neq \theta_0$  retains the null hypothesis if and only if the level  $(1 - \alpha) \times 100\%$  CI covers  $\theta_0$ .

# Assumptions

- ▶ We assume that:
  - ▶ The data are two SRSs from the 2 parent populations (e.g. moms with toxemia and moms without toxemia).
  - ▶ Also, either
    1. the data measurements in the parent populations are normally distributed with mean  $\mu_i$  and sd  $\sigma_i$ , written  $N(\mu_i, \sigma_i)$ , or
    2. the sample size  $n = n_1 + n_2$  is large enough to rely on the CLT for the sample means  $\bar{X}_1$  and  $\bar{X}_2$  being approximately normally distributed.
- ▶ Guidelines for  $n$  (*Basic Practice of Statistics* by D. Moore):
  - ▶ For  $n < 15$ , use the  $t$ -based CI and hypothesis test if the data look to be approximately normally distributed.
  - ▶ For  $15 \leq n < 40$ , use the  $t$ -based CI and hypothesis test, *except* in the presence of outliers or strong skewness in the data distribution.
  - ▶ For large samples ( $n \geq 40$ ), you can use the  $t$ -based CI and hypothesis test, even for clearly skewed distributions (because of the CLT).

# Checking the assumptions for the low-birth-weight example

- ▶ There were  $n = 100$  babies in this data set.
- ▶ According to the rules-of-thumb on the previous slide, we can use the  $t$ -based CI and hypothesis test even if the population distributions are skewed.

# Summary

- ▶ Inference for the difference between two population means is based on either  $Z$  (SDs known) or  $T$  (SDs unknown).
- ▶ Confidence intervals are of the form estimate  $\pm$  margin of error
  - ▶ the margin of error is a critical value times SE
- ▶ To test the null hypothesis  $H_0 : \mu_1 - \mu_2 = 0$  against an alternative  $H_a$  we compute a test statistic  $t$  (or  $z$  if SDs known) and  $p$ -value
  - ▶ can compare  $p$ -value to a significance level  $\alpha$  to obtain a test
- ▶ Inference is considered reliable when the parent populations are normal, or when rules of thumb for sample sizes are satisfied.