# Chapter 8: Comparing two independent populations

Part 1

https://dzwang91.github.io/stat324/

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

# Outline


1. Two sample t test with equal variance
2. Welch t test

# Two-sample t test

- We have two samples from two populations, label them 1 and 2. Let
  - $\mu_1$ = true mean of population 1
  - $\mu_2$ = true mean of population 2
  - $n_1$ = sample size taken from population 1
  - $n_2$ = sample size taken from population 2
  - $\sigma_1^2$ = true variance of population 1
  - $\sigma_2^2$ = true variance of population 2

# Two-sample t test

- We have two samples from two populations, label them 1 and 2. Let
  - $\mu_1$ = true mean of population 1
  - $\mu_2$ = true mean of population 2
  - $n_1$ = sample size taken from population 1
  - $n_2$ = sample size taken from population 2
  - $\sigma_1^2$ = true variance of population 1
  - $\sigma_2^2$ = true variance of population 2
- We wish to test:

$$H_0 : \mu_1 - \mu_2 = \delta \ vs. \ H_A : \mu_1 - \mu_2 \neq \delta$$

  or other one-tailed alternative hypothesis.
- An example: if we are interested in

$$H_0 : \mu_1 = \mu_2 \ vs. \ H_A : \mu_1 \neq \mu_2.$$

  It's equivalent to

$$H_0 : \mu_1 - \mu_2 = 0 \ vs. \ H_A : \mu_1 - \mu_2 \neq 0.$$

# Two-sample t test

- If we assume:
  - All of the data points are independent, both within and between populations
  - The two populations each follow normal distributions
  - The variances of the two populations are equal so that $\sigma_1^2 = \sigma_2^2 = \sigma^2$

  (how do we check these assumptions?)

  then the test statistic is:

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  Where:

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

- Compute the p-value or rejection region using $T$ distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom and then make a conclusion using given $\alpha$.

# Example

- The horned lizard Phrynosoma mcalli is named for the fringe of spikes around the back of the head. It was thought that the spikes may provide the lizard protection from its primary predator, the loggerhead shrike, Lanius ludovicanus, though there was not much existing quantitative evidence to support this.
- Researchers were interested in comparing two populations: the population of dead lizards known to be killed by shrikes, and the population of live lizards from the same geographic location. Random samples were taken from each population. The longest spike was measured on each sampled lizard.

# Example continued

- "Is there any difference in the size of the spikes between the two populations?"

- The data are as follows:
  - Dead Group: 17.65, 20.83, 24.59, 18.52, 21.40, 23.78, 20.36, 18.83, 21.83, 20.06
  - Live Group: 23.76, 21.17, 26.13, 20.18, 23.01, 24.84, 19.34, 24.94, 27.14, 25.87, 18.95, 22.61

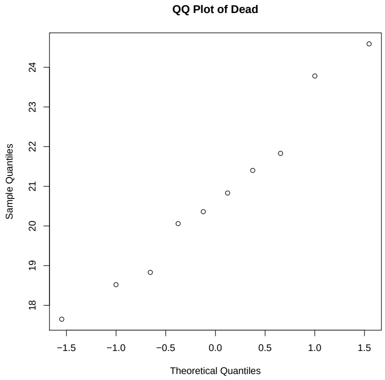
# Example continued

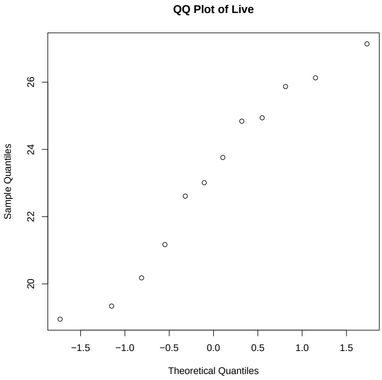
- "Is there any difference in the size of the spikes between the two populations?"
- The data are as follows:
  - Dead Group: 17.65, 20.83, 24.59, 18.52, 21.40, 23.78, 20.36, 18.83, 21.83, 20.06
  - Live Group: 23.76, 21.17, 26.13, 20.18, 23.01, 24.84, 19.34, 24.94, 27.14, 25.87, 18.95, 22.61
- Each data point is independent from the context.

# QQ plot

# QQ plot

- Separate QQ plots for each sample:

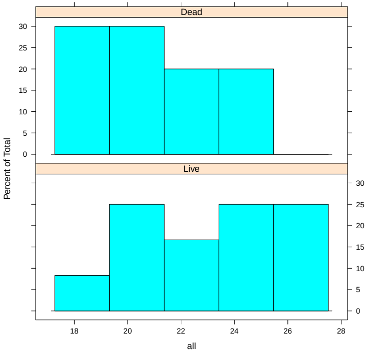



Each sample comes from a normal population.

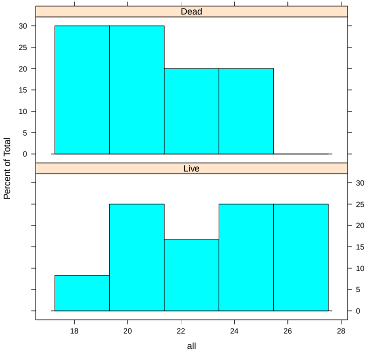
# Histogram

- Histograms from each group:



# Histogram

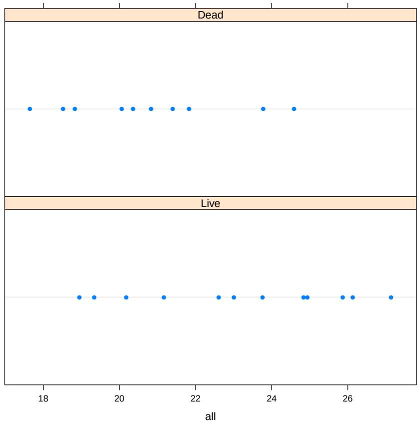
- Histograms from each group:




Even with the small sample size you can see that the live group seems to be shifted to the right a bit.

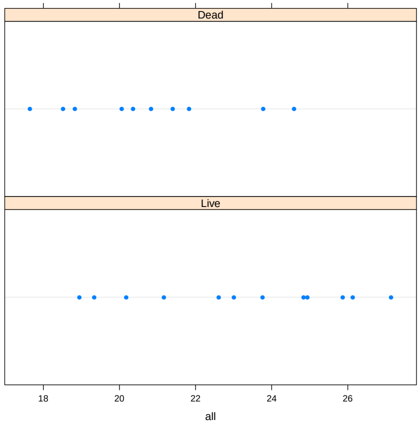
# Dotplot

- We call the plot of raw data a dotplot:



# Dotplot

- We call the plot of raw data a dotplot:




- Dotplot is good when there isn't much data, but when there's a lot, sometimes it's hard to see the important aspects of the data.

# Boxplot

- When there is a lot of data, a good choice for showing the rough location and spread of data is called a boxplot.
- To make a boxplot:
  - Plot a bar at the median, and at the first and third quartiles.
  - Connect the ends of the bars to make a box with a line in it.
  - Extend whiskers out to a maximum of 1.5*IQR up from the third quartile and down from the first quartile.
  - Any other data point outside of that range gets a dot.

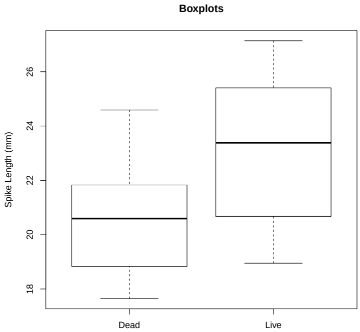
# Boxplot

- When there is a lot of data, a good choice for showing the rough location and spread of data is called a boxplot.
- To make a boxplot:
  - Plot a bar at the median, and at the first and third quartiles.
  - Connect the ends of the bars to make a box with a line in it.
  - Extend whiskers out to a maximum of 1.5*IQR up from the third quartile and down from the first quartile.
  - Any other data point outside of that range gets a dot.
- Here is some numerical summary information that will be helpful:

| Group | $n$ | Mean | Sample SD | 1st Q | Median | 3rd Q |
|---|---|---|---|---|---|---|
| Dead | 10 | 20.79 | 2.22 | 19.14 | 20.59 | 21.72 |
| Live | 12 | 23.16 | 2.76 | 20.92 | 23.16 | 25.17 |

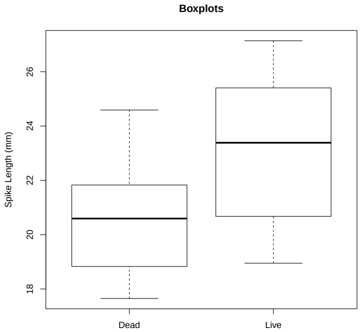
# Boxplot continued

- Boxplots side by side:



# Boxplot continued

- Boxplots side by side:




- In this boxplot, as we did in the histogram and dotplot, live seems generally higher. The spread also seems about the same.

# Back to example

- In this example, we don't know the population standard deviation.
- When the sample sizes are similar in the two groups, if $0.5 \leq \frac{s_1}{s_2} \leq 2.0$, then we can assume the population standard deviations are equal.
- In this case, $\frac{s_1}{s_2} = \frac{2.22}{2.76} = 0.8$, so we should be safe.

# Back to example

- In this example, we don't know the population standard deviation.
- When the sample sizes are similar in the two groups, if $0.5 \leq \frac{s_1}{s_2} \leq 2.0$, then we can assume the population standard deviations are equal.
- In this case, $\frac{s_1}{s_2} = \frac{2.22}{2.76} = 0.8$, so we should be safe.
- **Step 1:** let $\mu_{dead}$ = Mean of Dead Population, and $\mu_{live}$ = Mean of Live Population, then

$$H_0 : \mu_{dead} = \mu_{live} \text{ vs. } H_A : \mu_{dead} \neq \mu_{live}.$$

# Back to example

- In this example, we don't know the population standard deviation.
- When the sample sizes are similar in the two groups, if $0.5 \leq \frac{s_1}{s_2} \leq 2.0$, then we can assume the population standard deviations are equal.
- In this case, $\frac{s_1}{s_2} = \frac{2.22}{2.76} = 0.8$, so we should be safe.
- **Step 1:** let $\mu_{dead}$ = Mean of Dead Population, and $\mu_{live}$ = Mean of Live Population, then

$$H_0 : \mu_{dead} = \mu_{live} \text{ vs. } H_A : \mu_{dead} \neq \mu_{live}.$$

Or equivalently:

$$H_0 : \mu_{dead} - \mu_{live} = 0 \text{ vs. } H_A : \mu_{dead} - \mu_{live} \neq 0.$$

# Example continued

- **Step 2:**
  - Test statistic is

$$t = \frac{\bar{X}_{dead} - \bar{X}_{live} - 0}{S_p\sqrt{\frac{1}{n_{dead}} + \frac{1}{n_{live}}}},$$

    where

$$S_p^2 = \frac{(n_{dead}-1)S_{dead}^2 + (n_{live}-1)S_{live}^2}{n_{dead}+n_{live}-2}$$

  - In this example, $s_{dead}^2 = 2.22^2 = 4.93$, $s_{live}^2 = 2.76^2 = 7.62$, $s_p^2 = \frac{(10-1)4.93+(12-1)7.62}{10+12-2} = 6.41$, $T_{obs} = \frac{20.79-23.16-0}{\sqrt{6.41}\sqrt{\frac{1}{10}+\frac{1}{12}}} = -2.195$, and degrees of freedom $n_{dead} + n_{live} - 2 = 10 + 12 - 2 = 20$.

# Example continued

- **Step 3:** p-value$=2 \times P(T_{20} > 2.195) = 0.04$.
- **Step 4:** Given $\alpha = 0.05$, since the p-value is smaller than $\alpha$, we reject the null hypothesis.

**What if the equal variance assumption is not satisfied?**

# Outline

1. Two sample t test with equal variance

2. Welch t test

# Welch t test

Suppose we have independent simple random samples from normal populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$ where $\sigma_1 \neq \sigma_2$.

1. State hypotheses: $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$( or some other one-tailed $H_A$).
2. Check assumptions: independence, normality, unequal variance.
3. Find the test statistic: $t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$
4. Find the degrees of freedom:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

   and round down.
5. Find the p-value
6. Draw a conclusion.

# Example

- Concrete used for roadways or buildings is often reinforced with a material that is placed inside the setting concrete. A common example of this is called 'rebar' which is short for 'reinforcing bar' and is usually made out of steel. It is desirable that the reinforcing material is strong and corrosion resistant. Steel is strong, but tends to corrode over time, so experiments were conducted to test two corrosion resistant materials, one made of fiberglass and the other made of carbon.
- 8 beams with fiberglass reinforcement, and 11 beams with carbon reinforcement were poured, and each was then subjected to a load test, which measures the strength of the beam. Strength is measured in kN (kiloNewtons), which is a measure of the force required to break the beam.

# Example continued

- "Is there any difference in the strength of the two types of beams?"

$$H_0 : \mu_{fiber} = \mu_{carbon}, H_A : \mu_{fiber} \neq \mu_{carbon}$$

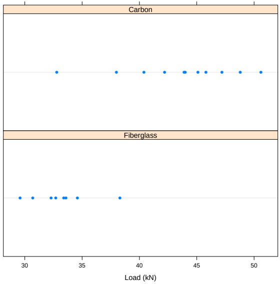
- The data are as follows:
    - Fiberglass: 38.3, 29.6, 33.4, 33.6, 30.7, 32.7, 34.6, 32.3
    - Carbon: 48.8, 38.0, 42.2, 45.1, 32.8, 47.2, 50.6, 44.0, 43.9, 40.4, 45.8

# Example continued

- Each data point is independent from the context.

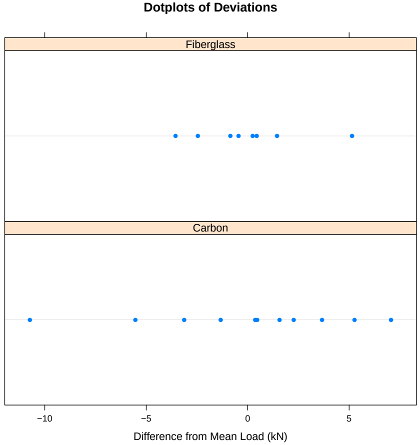
# Example continued

- Each data point is independent from the context.
- By dotplot, the mean for carbon looks a bit higher, but the sd is also larger.



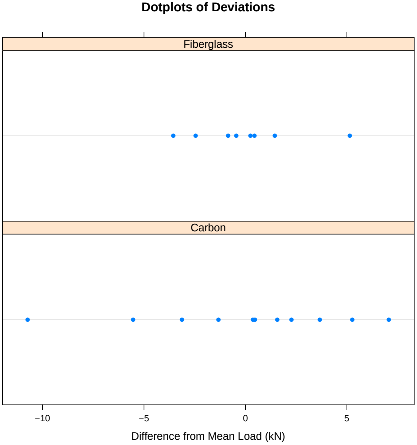
# Example continued

- We can also look at differences from means:



# Example continued

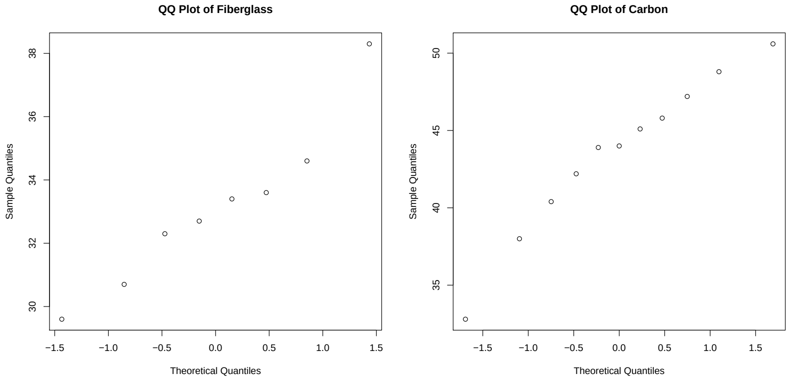
- We can also look at differences from means:




The constant variance assumption doesn't hold.

# Example continued

- Separate QQ plots for each sample:




Each sample comes from a normal population.

# Example continued

- Therefore we use the Welch t test:
  - The test statistic is

$$T = \frac{\bar{X}_{fiber} - \bar{X}_{carbon}}{\sqrt{\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}}}$$

# Example continued

- Therefore we use the Welch t test:
  - The test statistic is
$$T = \frac{\bar{X}_{fiber} - \bar{X}_{carbon}}{\sqrt{\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}}}$$
  - The distribution of this statistic under $H_0$ is only approximately $T$ and the degrees of freedom are
$$\nu = \frac{\left(\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}\right)^2}{\frac{(s^2_{fiber}/n_{fiber})^2}{n_{fiber}-1} + \frac{(s^2_{carbon}/n_{carbon})^2}{n_{carbon}-1}}$$

# Example continued

- In the example:
  - $T_{obs} = \frac{33.15 - 43.53 - 0}{\sqrt{\frac{2.63^2}{8} + \frac{5.06^2}{11}}} = -5.81$
  - $\nu = \frac{\left(\frac{2.63^2}{8} + \frac{5.06^2}{11}\right)^2}{\frac{(2.63^2/8)^2}{8-1} + \frac{(5.06^2/11)^2}{11-1}} = 15.7$, round down to 15
- The p-value is $2 \times P(T_{15} > 5.81) < 0.001$, so there is evidence that the two kinds of materials are not equally strong. It seems the carbon is stronger and would be preferred.

# Guideline

- In general, the consequences of the choice of using equal or unequal variances can be summarized in two statements:

# Guideline

- In general, the consequences of the choice of using equal or unequal variances can be summarized in two statements:
  - If the variances are truly equal, but are assumed to be unequal, the test loses some power, but is still a good test.

# Guideline

- In general, the consequences of the choice of using equal or unequal variances can be summarized in two statements:
    - If the variances are truly equal, but are assumed to be unequal, the test loses some power, but is still a good test.
    - If the variances are truly different, but they are assumed to be equal, the test can make wildly incorrect conclusions.

# Guideline

- In general, the consequences of the choice of using equal or unequal variances can be summarized in two statements:
  - If the variances are truly equal, but are assumed to be unequal, the test loses some power, but is still a good test.
  - If the variances are truly different, but they are assumed to be equal, the test can make wildly incorrect conclusions.
- Therefore, if there is any doubt about the equality of the variances, then use Welch t test.

# What's the next?

We'll discuss how to compare two population proportions in next lecture.