Chapter 8: Comparing two independent populations

Duzhe Wang

Part 2: The Welch T-test

https://dzwang91.github.io/stat371/

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

- The primary research question was, "Is there any difference in the strength of the two types of beams?" Thus we wish to test:

$$H_0 : \mu_{fiber} - \mu_{carbon} = 0, H_A : \mu_{fiber} - \mu_{carbon} \neq 0$$

- 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.
- Numerical summaries:

Beam Type	Sample Size	Mean	SD
Fiber	8	33.15	2.63
Carbon	11	43.53	5.06

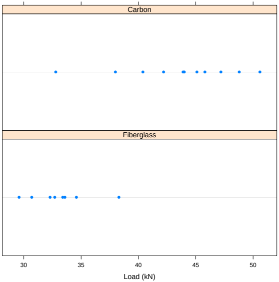
$\frac{s_1}{s_2} = \frac{2.63}{5.06} = 0.52$, so while still strictly within the allowed limits, it is very close to the edge.

Example continued

- Each data point is independent from the context.
- Numerical summaries:

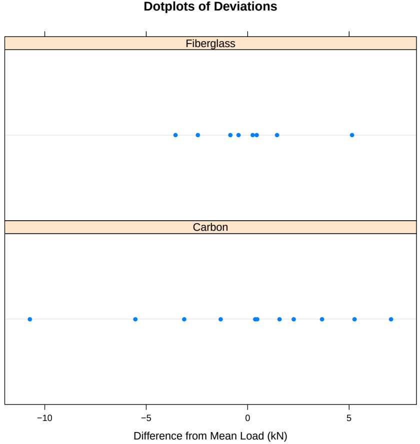
Beam Type	Sample Size	Mean	SD
Fiber	8	33.15	2.63
Carbon	11	43.53	5.06

  $\frac{s_1}{s_2} = \frac{2.63}{5.06} = 0.52$, so while still strictly within the allowed limits, it is very close to the edge.
- By dotplot, the mean for carbon looks a bit higher, but the sd is also larger.



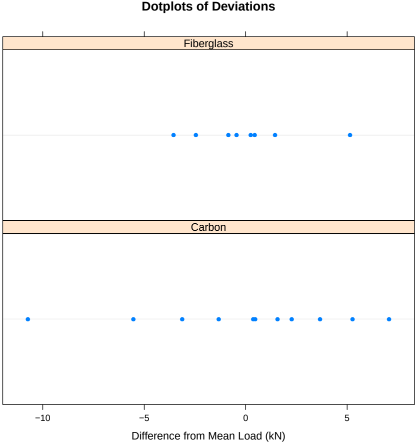
Example continued

- We can also look at differences from means.



Example continued

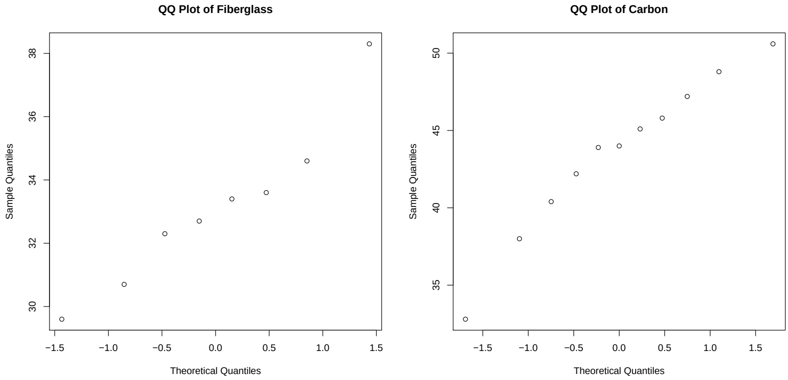
- We can also look at differences from means.



The constant variance assumption doesn't seem to be met, because there appears to be more spread in the carbon group. To be safe, we may not want to assume equal variances.

Example continued

- Separate QQ plots for each sample:



Each sample comes from a normal population.

Example continued

- Therefore, Instead of using two sample T-test, we use the Welch T-test.

Example continued

- Therefore, Instead of using two sample T-test, we use the Welch T-test.
- The test statistic is

$$T = \frac{\bar{X}_{fiber} - \bar{X}_{carbon} - 0}{\sqrt{\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}}}$$

Example continued

- Therefore, Instead of using two sample T-test, we use the Welch T-test.
- The test statistic is

$$T = \frac{\bar{X}_{fiber} - \bar{X}_{carbon} - 0}{\sqrt{\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}}}$$

  - The distribution of this statistic 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

- Therefore, Instead of using two sample T-test, we use the Welch T-test.
- The test statistic is

$$T = \frac{\bar{X}_{fiber} - \bar{X}_{carbon} - 0}{\sqrt{\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}}}$$

  - The distribution of this statistic is only approximately T and the degrees of freedom are

$$\nu = \frac{\left(\frac{s^2_{fiber}}{n_{fiber}} + \frac{s^2_{carbon}}{n_{carbon}}\right)^2}{\frac{(s^2_{fiber}/n_{fiber})^2}{n_{fiber}-1} + \frac{(s^2_{carbon}/n_{carbon})^2}{n_{carbon}-1}}$$

  - It is likely that the degrees of freedom will not be a whole number. It is common to round down to the nearest integer.

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.

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 where $\sigma_1 \neq \sigma_2$.

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_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).

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_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.

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_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: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_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.

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_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

Recap of Welch T-test

Suppose we have independent simple random samples from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_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.

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, it's generally safer to allow them to differ.

What's the next?

We'll discuss how to compare two population proportions in the third part of Chapter 8.