Chapter 8: Comparing Two Independent Populations
Part 2: The Welch T-test(Normal Data with Unequal Variances)

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.

The primary research question was, "Is there any difference in the strength of the two types of beams?" Since neither of the materials had been used before, there was no prior hypothesis about which one might be better. Thus we wish to test:

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

Here, we are experimentally manipulating the conditions, so we will be able to infer causation. The data are as follows:

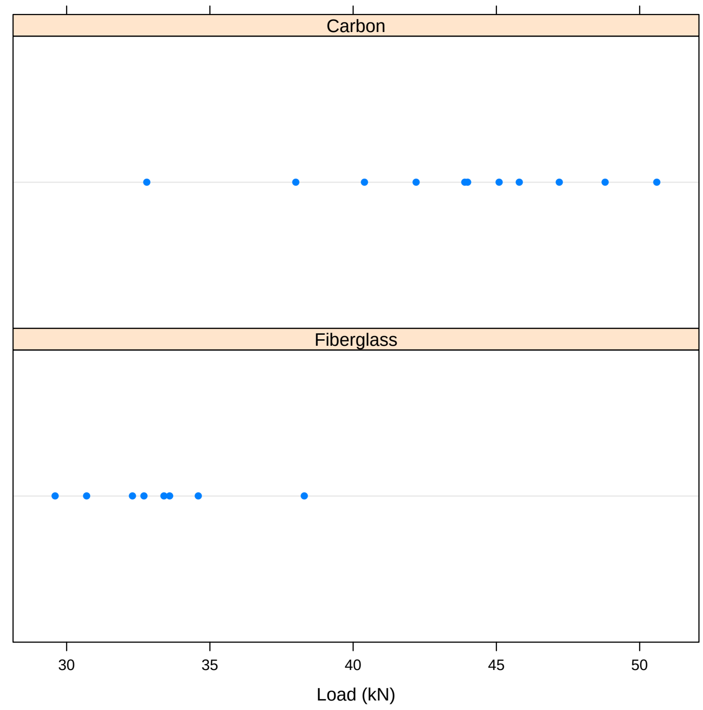
$$Fiberglass : 38.3, 29.6, 33.4, 33.6, 30.7, 32.7, 34.6, 32.3$$

$$Carbon : 48.8, 38.0, 42.2, 45.1, 32.8, 47.2, 50.6, 44.0, 43.9, 40.4, 45.8$$

As always, we proceed by first computing graphical and numerical summaries. In the interest of time, we will stick to the most useful ones here:

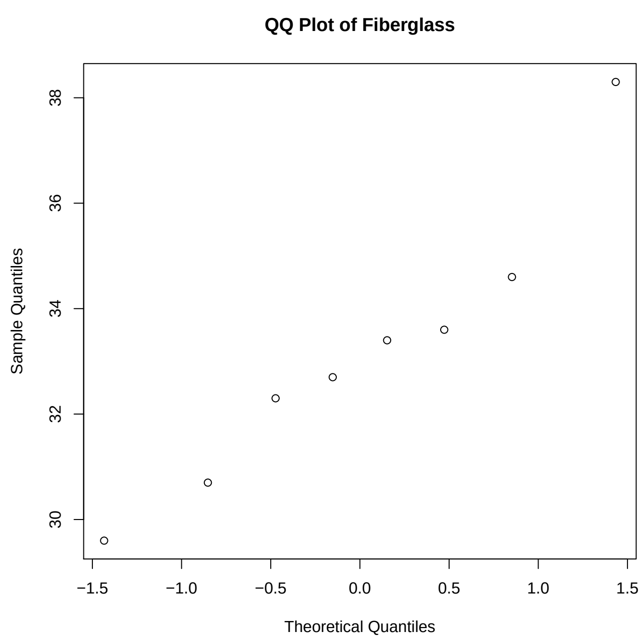
Beam Type	**Sample Size**	**Mean**	**SD**
Fiber	8	33.15	2.63
Carbon	11	43.53	5.06

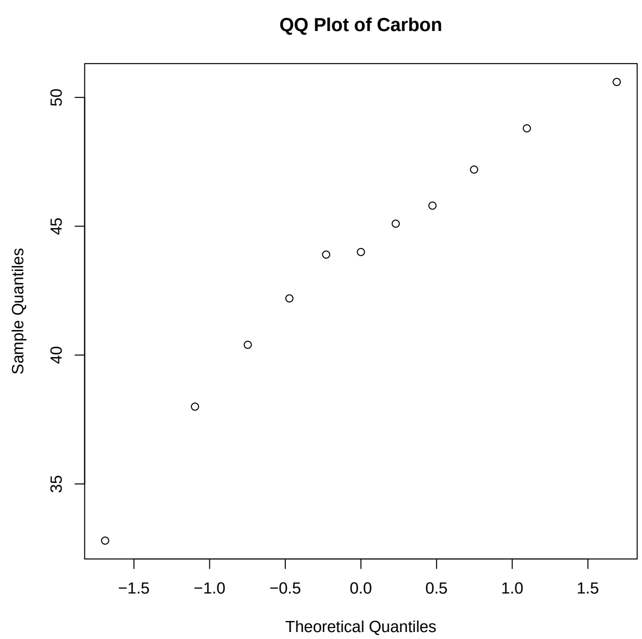
The mean for carbon looks a bit higher, but the sd is also larger. This is confirmed by a dotplot:



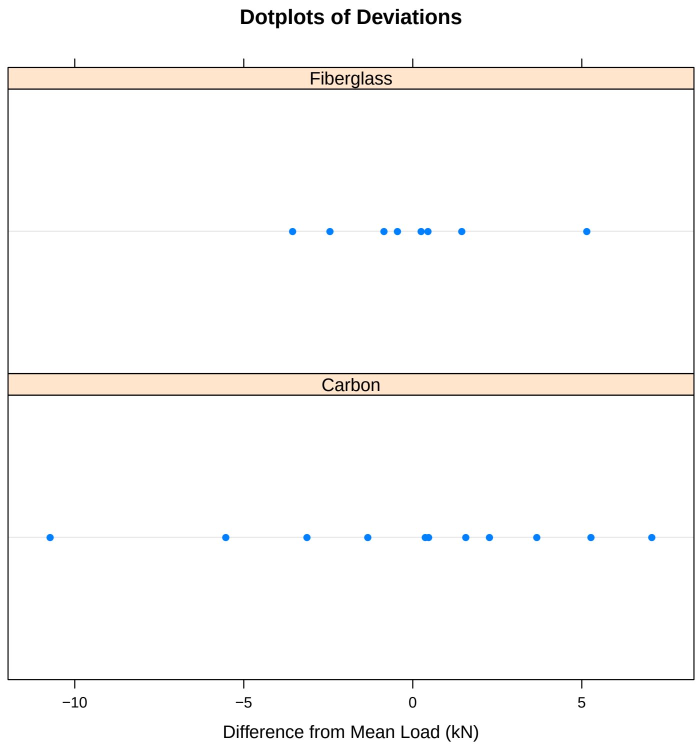
The graph confirms what we saw in the numerical summaries. The carbon seems to be a bit higher, but also more variable.

Can we use the same test for this data as we did for the lizard data? What were the assumptions of the previous test? Independence, Normality, Constant Variance. Independence is probably fine from the context. QQ plots will help us evaluate normality:





From the separate QQ plots, normality looks fine. What about constant variance? We can again 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. Checking our guideline, $\frac{s_1}{s_2} = \frac{2.63}{5.06} = 0.52$, so while still strictly within the allowed limits, it is very close to the edge, and the sample sizes do differ. To be safe, we may not want to assume equal variances.

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 not assumed to be equal, the test loses some power, but is still a good test.
- If the variances are truly different, but they are assumed equal, the test can make wildly incorrect conclusions.

There is usually more to lose in the second case. Therefore, if there is any doubt about the equality of the variances, it's generally safer to allow them to differ.

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

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

It turns out that the distribution of this statistic is only approximately T, and the degrees of freedom are nasty:

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

It is likely that the degrees of freedom will not be a whole number. It is possible to find T probabilities for fractional degrees of freedom using a computer, but it is also common to round down to the nearest integer, especially when you need to use a table.

For the data in the above example:

- $t = \frac{33.15-43.53-0}{\sqrt{\frac{2.63^2}{8}+\frac{5.06^2}{11}}} = -5.80967$
- $\nu = \frac{\left(\frac{2.63^2}{8}+\frac{5.06^2}{11}\right)^2}{\frac{(2.63^2/8)^2}{8-1}+\frac{(5.06^2/11)^2}{11-1}} = 15.7$, round down to 15

The p-value is < 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: 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 $H_A : \mu_1 \neq \mu_2$.
2. Check assumptions.
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: p-value$=2 * P(T_\nu > |t_{obs}|)$.
6. Draw a conclusion.