TODAY

- Finish t-test (paired, 2-sample)
- Simple linear regression

PAIRED t-TEST

See https://www.stat.purdue.edu/~tqin/system101/method/method_pair_t_sas.htm

Assumptions:

1. The two groups of data are dependent:
    - One group: each individual's change over time.
    - Difference between two related groups (siblings, partners, twins, left-tire vs right-tire, etc.)
2. The differences between control and treatment follow normal distribution;

In SAS:

```
proc ttest data=response sides=2 alpha=0.05 h0=0;
    title "Paired sample t-test example";
    paired Control * Treatment;
run;
```

The data (from Purdue course):

Total rows: 6 Total columns: 2 Rows 1-6

	Control	Treatment
1	90	95
2	87	92
3	100	104
4	80	89
5	95	101
6	90	105

Paired sample t-test example

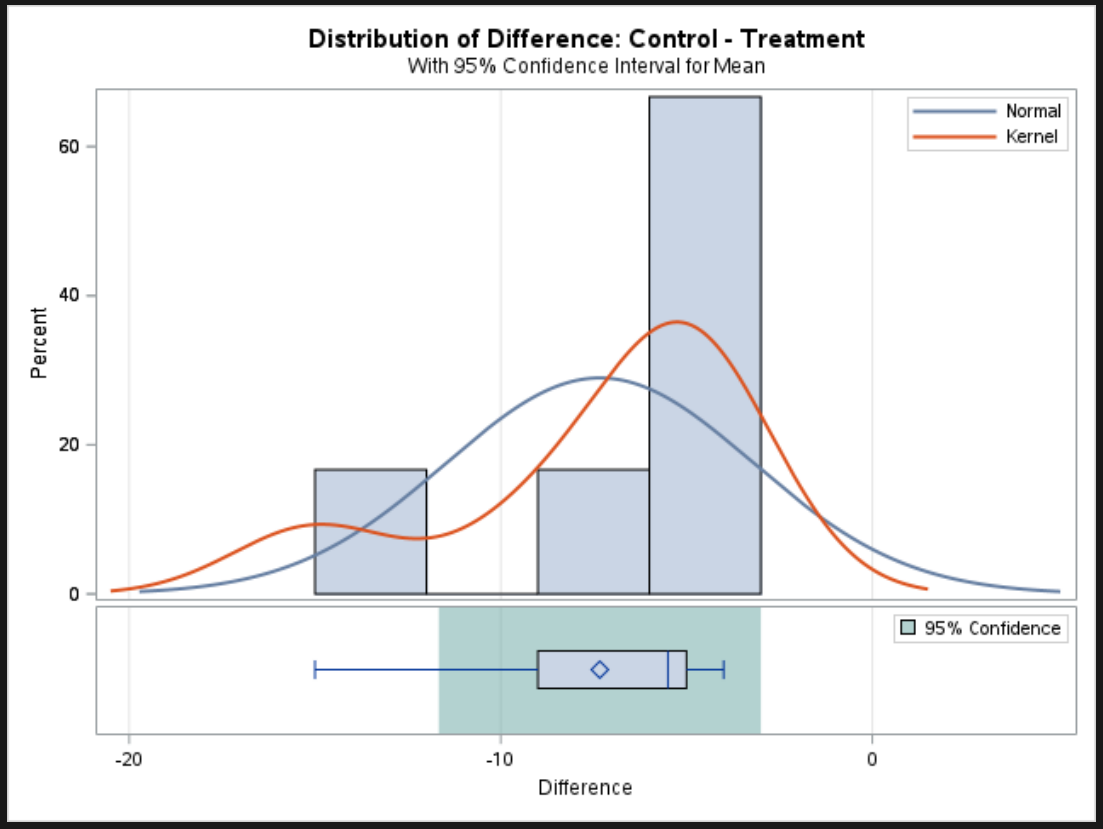
The TTEST Procedure

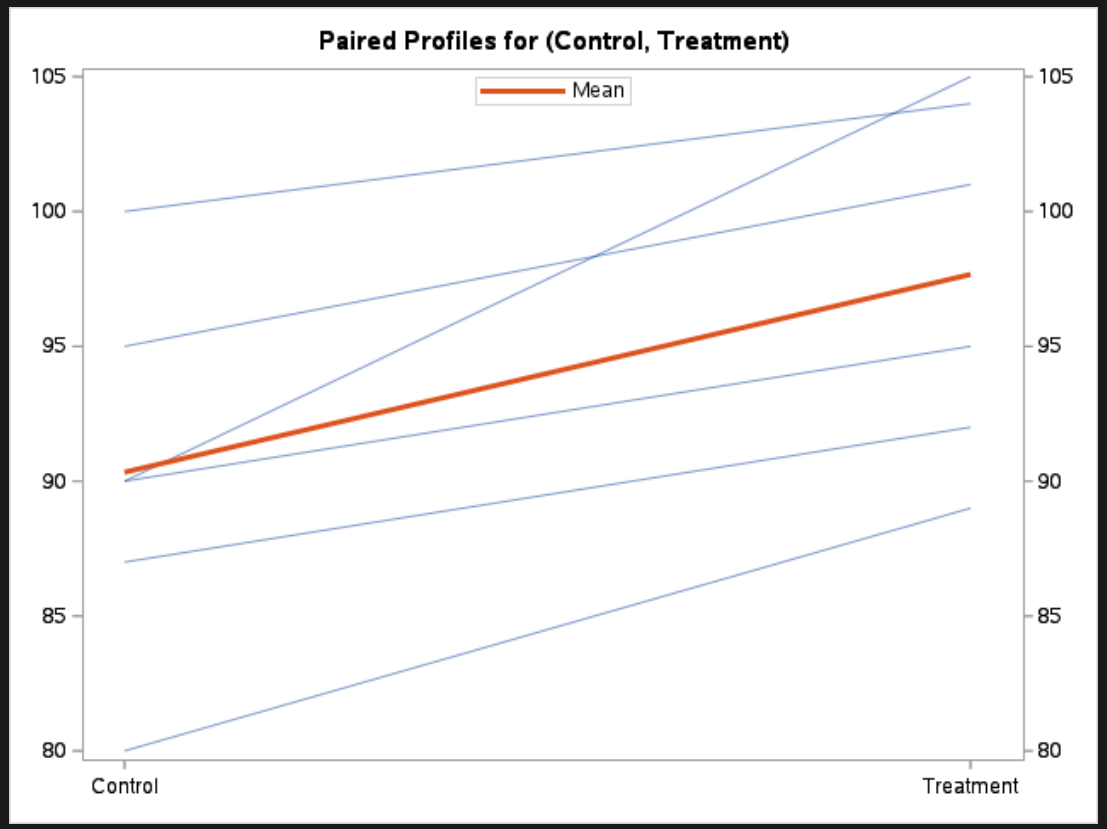
Difference: Control - Treatment

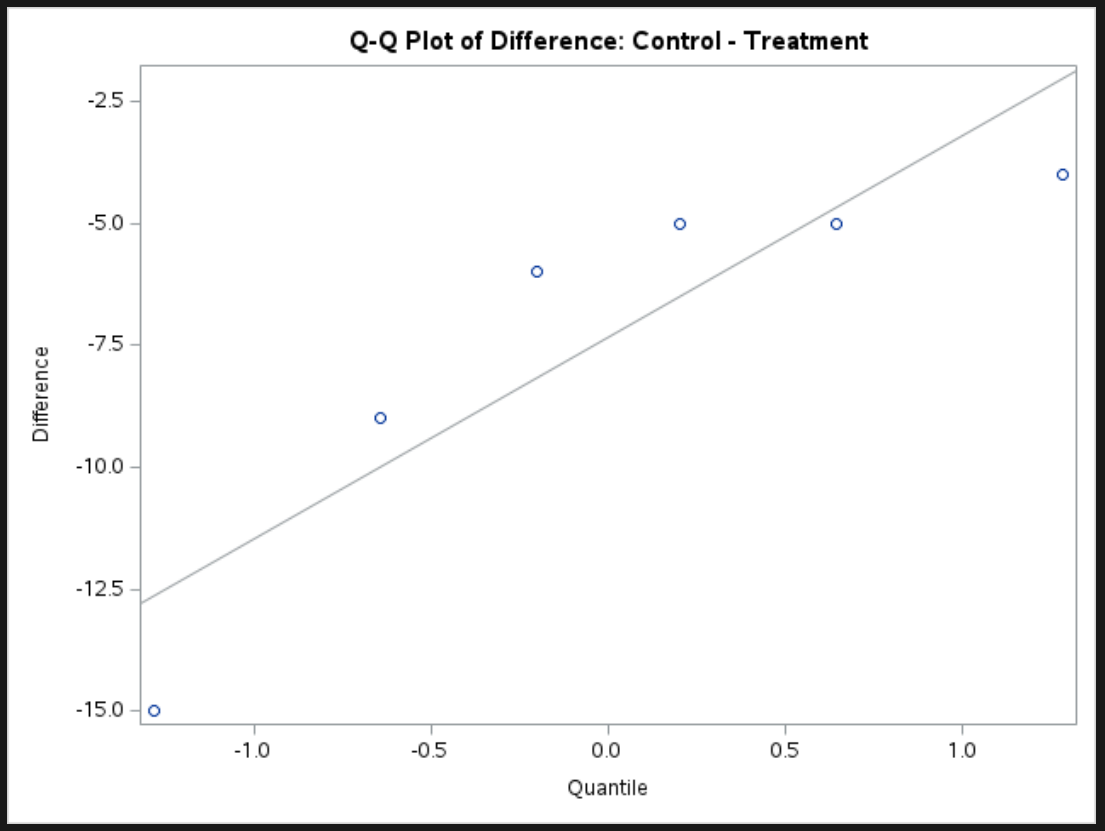
N	Mean	Std Dev	Std Err	Minimum	Maximum
6	-7.3333	4.1312	1.6865	-15.0000	-4.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-7.3333	-11.6687	-2.9979	4.1312	2.5787	10.1322

DF	t Value	Pr > |t|
5	-4.35	0.0074


Distribution of Difference: Control - Treatment
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
60
40
20
0
95% Confidence
-20
-10
0
Difference


Paired Profiles for (Control, Treatment)
Mean
105
100
95
90
85
80
Control
Treatment


Q-Q Plot of Difference: Control - Treatment
Difference
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-1.0
-0.5
0.0
0.5
1.0
Quantile

2-SAMPLE t-TEST

See https://www.stat.purdue.edu/~tqin/system101/method/method_two_t_sas.htm

Assumptions:

1. There is one continuous dependent variable and one categorical independent variable (with 2 levels);
2. The two samples are independent;
3. The two samples follow normal distributions.

Here the test statistics for $H_0 : \mu_1 - \mu_2 = 0$ is:

$$t = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

which follows an approximate t-distribution with df given by a more complicated formula (Welch–Satterthwaite equation) for the Welch's unequal variance t-test.

In SAS:

```
filename myurl url "https://www.stat.purdue.edu/~tqin/system10
proc import
    datafile=myurl out=read dbms=csv replace;
    getnames = yes;
run;

/* Do t-test. CLASS variable defines the groups. */
proc ttest data=read sides=2 alpha=0.05 h0=0;
     title "Two sample t-test example";
     class method;
    var grade;
run;
```

Total rows: 10 Total columns: 2 Rows 1-10

	Method	Grade
1	c	80
2	c	93
3	c	83
4	c	89
5	c	98
6	t	100
7	t	103
8	t	104
9	t	99
10	t	102

Two sample t-test example

The TTEST Procedure

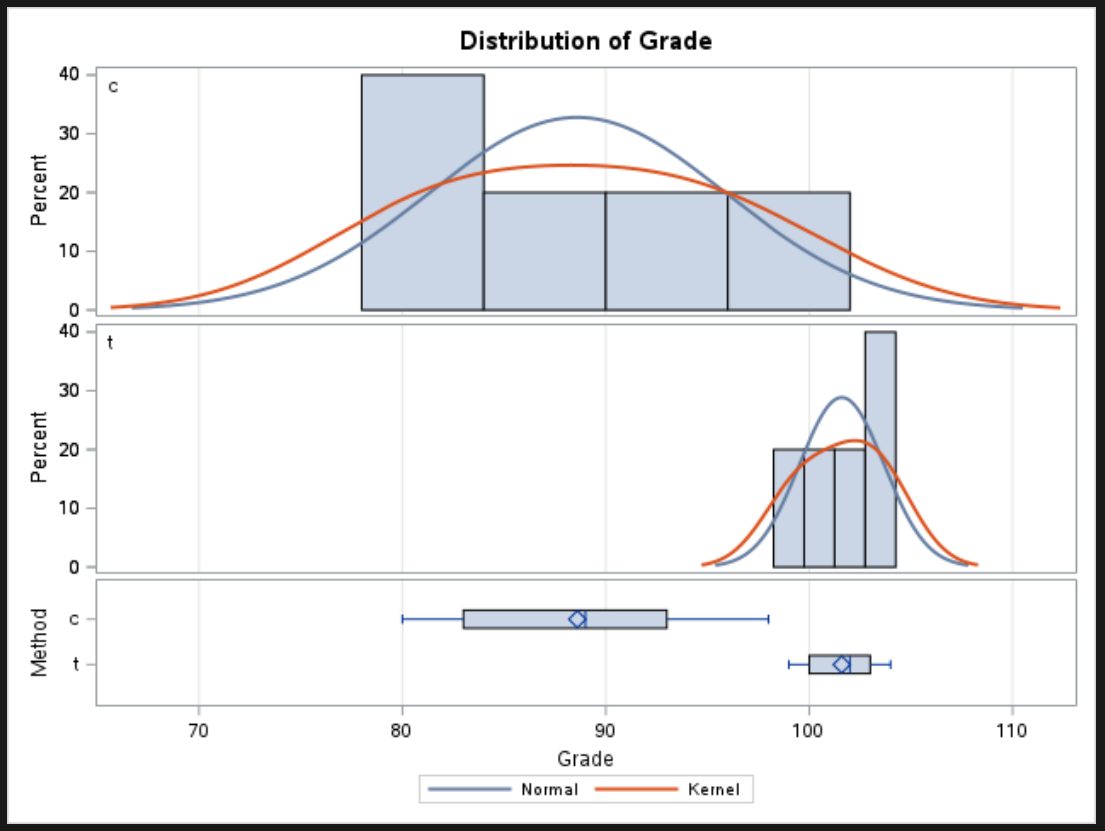
Variable: Grade

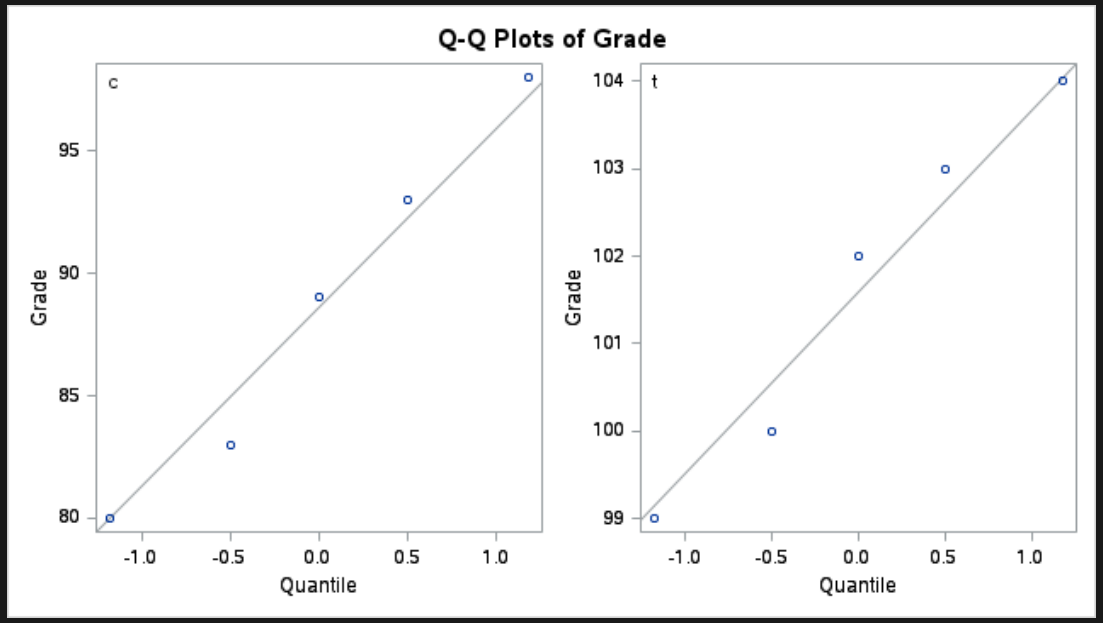
Method	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
c		5	88.6000	7.3007	3.2650	80.0000	98.0000
t		5	101.6	2.0736	0.9274	99.0000	104.0
Diff (1-2)	Pooled		-13.0000	5.3666	3.3941		
Diff (1-2)	Satterthwaite		-13.0000		3.3941		

Method	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
c		88.6000	79.5350	97.6650	7.3007	4.3741	20.9789
t		101.6	99.0252	104.2	2.0736	1.2424	5.9587
Diff (1-2)	Pooled	-13.0000	-20.8268	-5.1732	5.3666	3.6249	10.2811
Diff (1-2)	Satterthwaite	-13.0000	-21.9317	-4.0683			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8	-3.83	0.0050
Satterthwaite	Unequal	4.6412	-3.83	0.0141

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	4	4	12.40	0.0318


Distribution of Grade
c
t
Percent
40
30
20
10
0
Method
c
t
70
80
90
100
110
Grade
Normal
Kernel


Q-Q Plots of Grade
c
t
Grade
Quantile
95
90
85
80
104
103
102
101
100
99
-1.0
-0.5
0.0
0.5
1.0