Section 7.1 - Inference on the Mean of a Population

Statistics 104

Autumn 2004

t Distributions

Our initial approach to inference on the mean of a population made the assumption that the variance of the population being sampled from was known.

Most of the time this is an unreasonable assumption and one that is not required to be made.

As mentioned many times before, σ can be estimated by s, so we can use this to get the standard error

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Instead of basing inference on

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

it will be based on

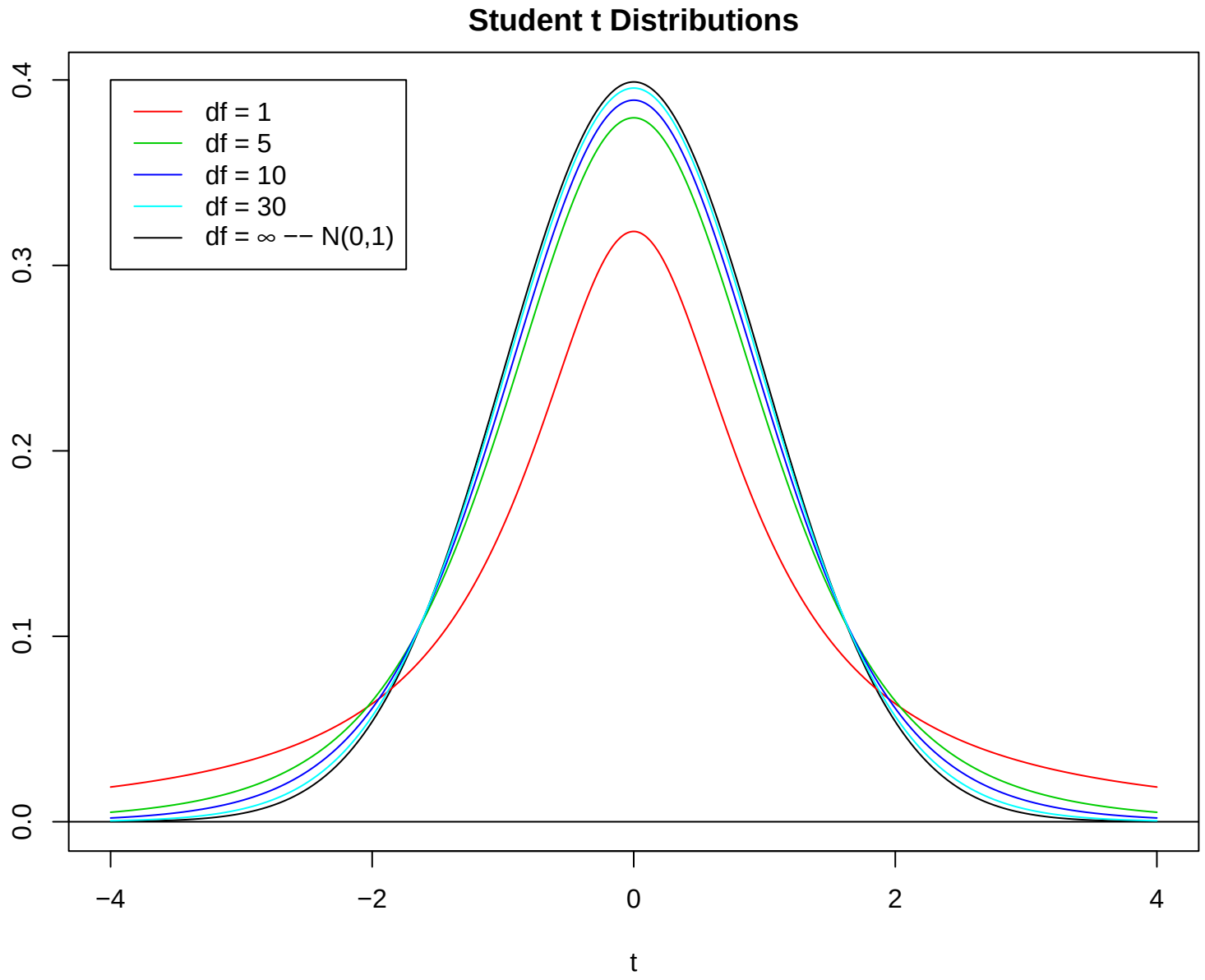
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

William Sealey Gosset
(aka Student)

If the observations $X_1, X_2, \ldots, X_n$ are drawn from a SRS from a $N(\mu, \sigma)$ distribution, the quantity t has a Student t distribution with $n-1$ degrees of freedom (denoted by $t(n-1)$).

So for every sample size n, we get a different distribution.

The t distributions are centered at 0 and symmetric like the $N(0,1)$ distribution, but they are more spread out. As the degrees of freedom increases, the distribution gets closer and closer to a $N(0,1)$.


Student t Distributions
df = 1
df = 5
df = 10
df = 30
df = ∞ −− N(0,1)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
t

The extra variability in the t distributions makes sense. Since we are plugging in an estimate of σ, we are less certain about the distribution we are sampling from, so that extra uncertainty needs to be accounted for.

As we have more data, we have more information about the distribution we are sampling from, so our inferences should act more like the case when we know σ.

Confidence Intervals for a Population Mean

Similar to before, replace the normal critical value z^* with the t critical value t^*.

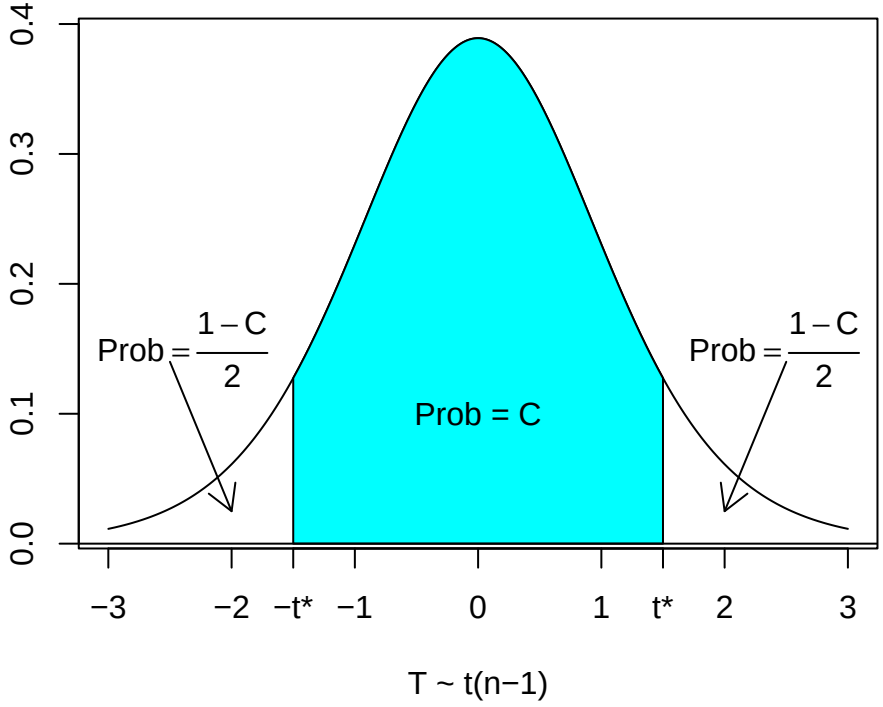
A $100C\%$ CI for μ is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* satisfies

$$P[-t^* \leq T \leq t^*] = C$$

and $T \sim t(n-1)$. The t critical values t^* can be obtained from Table D.



This interval is exact if the population distribution is normal and approximately correct for large n in other cases.

To get the correct critical value, go to row $n - 1$ and the column with confidence level C.

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

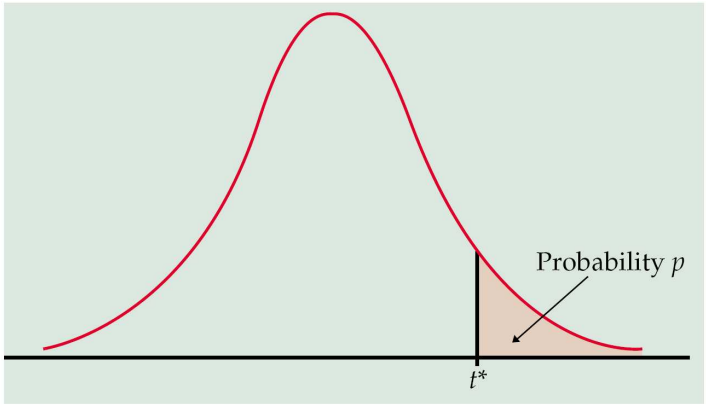
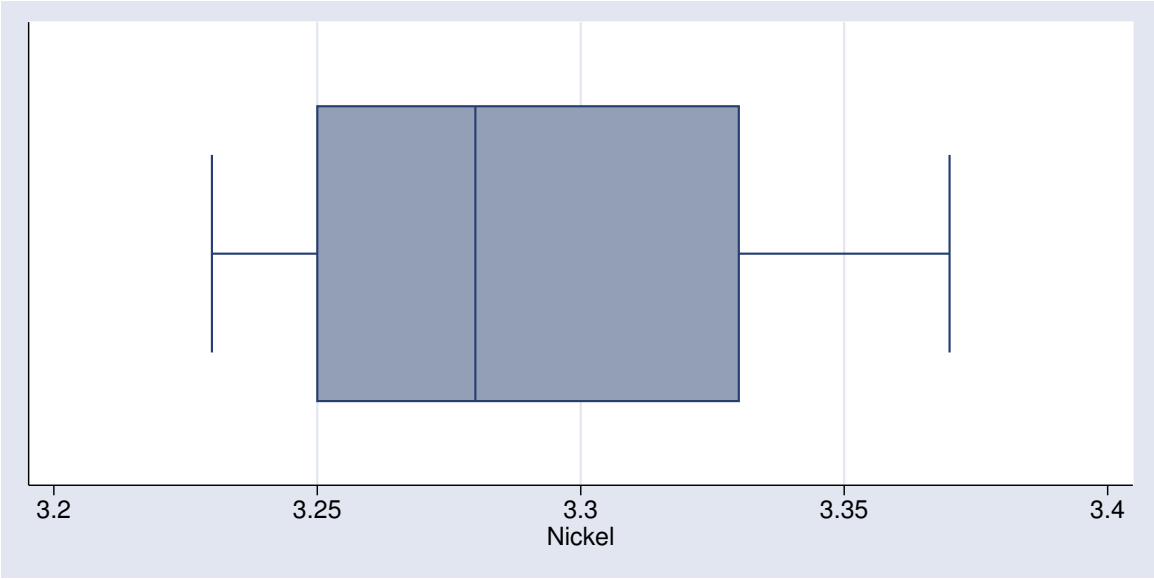


TABLE D t distribution critical values

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Example: Nickel content of ore

A new batch of ore is to be tested for its nickel content to determine whether it is consistent with the usual mean content of 3.25% that has been found in past batches. Ten sample were taken.



```
. summarize Nickel, detail

                          Nickel
-------------------------------------------------------------
      Percentiles      Smallest
 1%         3.23           3.23
 5%         3.23           3.24
10%        3.235           3.25       Obs                  10
25%         3.25           3.26       Sum of Wgt.          10

50%         3.28                      Mean                3.289
                        Largest       Std. Dev.       .0470106
75%         3.33           3.31
90%        3.355           3.33       Variance          .00221
95%         3.37           3.34       Skewness        .3705281
99%         3.37           3.37       Kurtosis        1.867116
```

Lets look at a 90% CI for μ. With $n-1=9, t^*=1.833$

$$\begin{aligned} & 3.289 \pm 1.833 \frac{0.0470}{\sqrt{10}} \\ = \quad & 3.289 \pm 1.833 \times 0.0148 \\ = \quad & 3.289 \pm 0.027 = (3.262, 3.316) \end{aligned}$$

```
. ci Nickel, level(90)

Variable |        Obs        Mean    Std. Err.       [90% Conf. Interval]
---------+--------------------------------------------------------------
  Nickel |         10       3.289    .0148661        3.261749    3.316251
```

One-sample t Test of a Population Mean

Again its similar to the normal z test. The test statistic for examing $H_0 : \mu = \mu_0$ is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

The p-values for this test statistic are

$$\begin{aligned} H_A : \mu > \mu_0 \qquad & p\text{-value} = P[T \geq t_{obs}] \\ H_A : \mu < \mu_0 \qquad & p\text{-value} = P[T \leq t_{obs}] \\ H_A : \mu \neq \mu_0 \qquad & p\text{-value} = 2 \times P[T \geq |t_{obs}|] \end{aligned}$$

where $T \sim t(n-1)$. These are analogous to the z test with the normal distribution there replaced by the appropriate t distribution.

For the nickel example with $H_0 : \mu = 3.25$,

$$t = \frac{3.289 - 3.250}{0.0470/\sqrt{10}} = \frac{0.039}{0.0149} = 2.6234$$

For $H_A : \mu \neq 3.25$,

$$p - \text{value} = 2 \times P[T \geq 2.6234] = 2 \times 0.0138 = 0.0277$$

So there appears to be some evidence that the mean nickel content isn't 3.25%.

```
. ttest Nickel == 3.25

One-sample t test

------------------------------------------------------------------------------
Variable |  Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  Nickel |   10       3.289    .0148661    .0470106    3.255371    3.322629
------------------------------------------------------------------------------
Degrees of freedom: 9

                     Ho: mean(Nickel) = 3.25

    Ha: mean < 3.25               Ha: mean != 3.25                Ha: mean > 3.25
       t =   2.6234                    t =   2.6234                    t =   2.6234
   P < t =   0.9862               P > |t| =   0.0277               P > t =   0.0138
```

How to get p-values without a computer

There are no tables similar to Table A, the CDF function for the $N(0,1)$ distribution, so it is not possible to get an exact p-valuewith a computer or calculator.

The reason for the lack of table is that you would need one for each possible sample size.

However, using Table D, we can put bounds on the p-value.

TABLE D t distribution critical values

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587

Find value in the row with $n-1$ df that flank $|t_{obs}| = 2.6234$.

In this case its 2.389 ($p_u = 0.02$) and 2.821 ($p_l = 0.01$).

Then for the possible three alternative hypotheses,

$$H_A : \mu > \mu_0 \qquad p_l \leq p\text{-value} \leq p_u$$
$$H_A : \mu < \mu_0 \qquad (1-p_u) \leq p\text{-value} \leq (1-p_l)$$
$$H_A : \mu \neq \mu_0 \qquad 2p_l \leq p\text{-value} \leq 2p_u$$

So for the two sided example, the bounds on the p-valueare 0.02 and 0.04. The true p-valueis 0.027, which is in the range.

What if the degrees of freedom you're interested in aren't in the table (for CIs and tests)?

TABLE D t distribution critical values

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Use the row above where the desired row should be. For example, if the desired df $= 45$, use the row for df $= 40$.

For CIs, this will give slightly wider CIs than if you use correct degrees of freedom, since the critical value is slightly larger with smaller dfs.

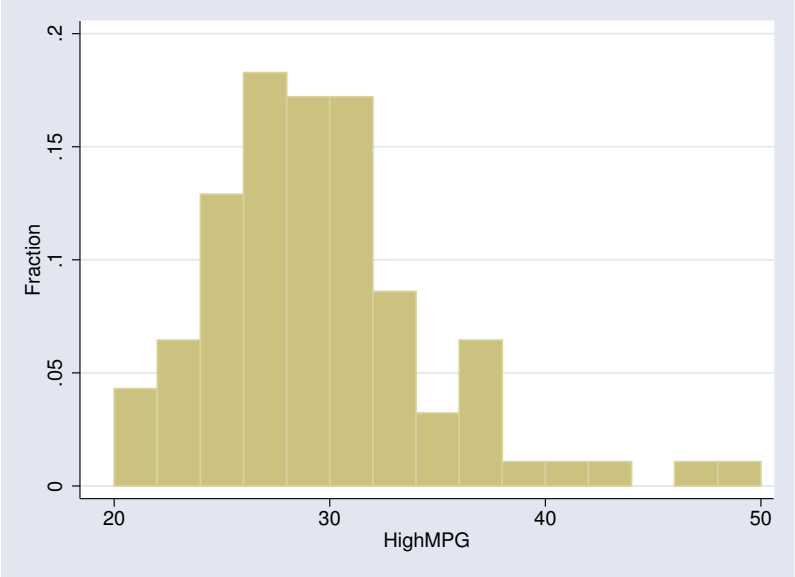
For 95% confidence

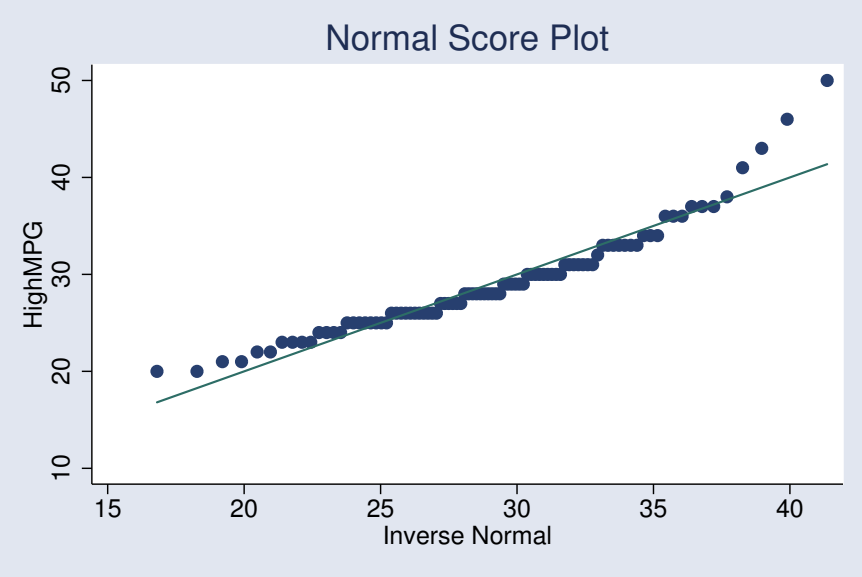
$$\text{df} = 40 \quad t^* = 2.021$$
$$\text{df} = 45 \quad t^* = 2.014$$

Similarly for tests, you will get slightly bigger p-values using this approximation.

When the degrees of freedom is smaller, you will see bigger difference between the rows. This approximation is more conservative for smaller sample sizes.

Look at EPA Highway MPG ratings





Let's suppose this was a sample from the population of 1993 cars models.

Suppose we were interested in testing $H_0 : \mu = 31$

```
. ttest HighMPG == 31

One-sample t test

------------------------------------------------------------------------------
Variable |   Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 HighMPG |    93    29.08602    .5528742    5.331726    27.98797    30.18408
------------------------------------------------------------------------------
Degrees of freedom: 92

                     Ho: mean(HighMPG) = 31

     Ha: mean < 31               Ha: mean != 31                Ha: mean > 31
       t =  -3.4619                    t =  -3.4619                  t =  -3.4619
   P < t =   0.0004              P > |t| =   0.0008              P > t =   0.9996
```

The normality assumption doesn't seem to be particularly reasonable here. The histogram and normal score plot both exhibit right skewness.

However the t test and t CI are fairly robust against non-normality except in the case of outliers and strong skewness.

Robust Procedure

> A statistical inference procedure is called **robust** if the probability calculations required are insensitive to violations of the assumptions made.

Since the t procedures are fairly robust, I wouldn't worry about the concluding that the mean highway MPG is not 31 MPG since the p-value is way below 0.05 for the two-sided test.

If instead, it had been say 0.047, you might want to be careful about making strong statements about the result being statistically significant.

Paired Comparison Design

Usually in statistics, we aren't interesting in one group of treatment, but in comparing groups or treatments.

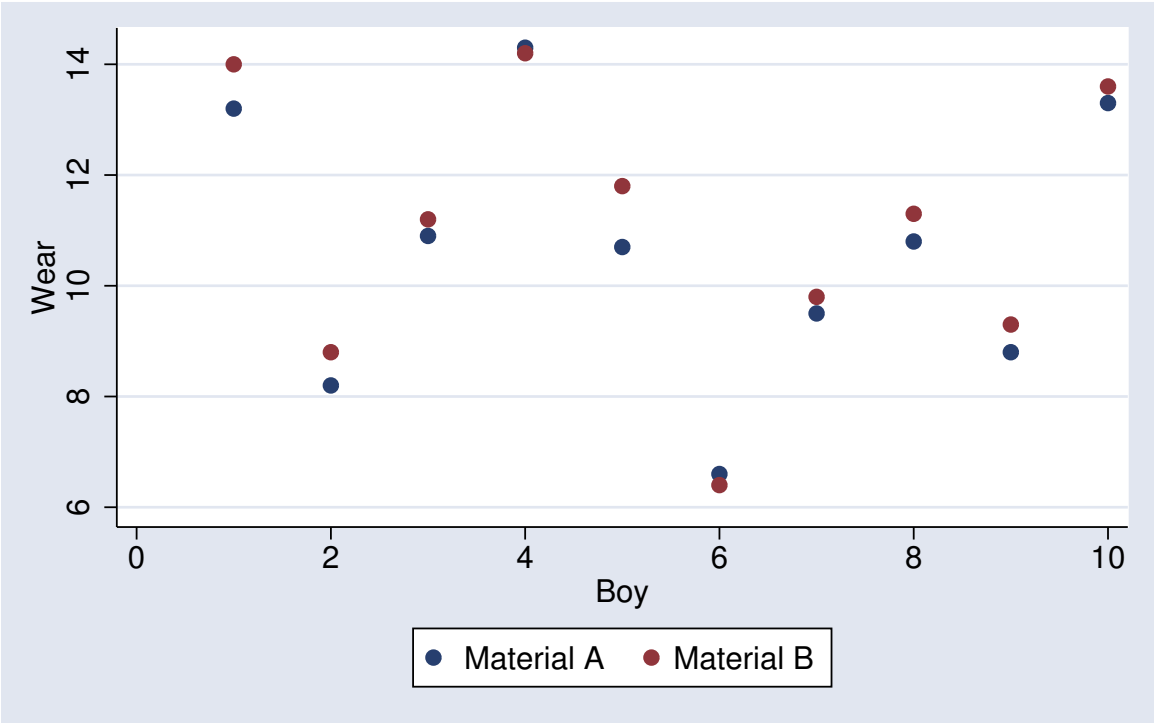
The paired comparison design is one of the easiest ways of comparing two treatments.

In this design, observations are paired, with one getting one treatment and the other getting the other treatment. This matching has to be done on the basis of the units and the design, not the data. Possible matching could be done with twins, right and left eye, etc.

Paired t Procedures

Example: Shoe sole material

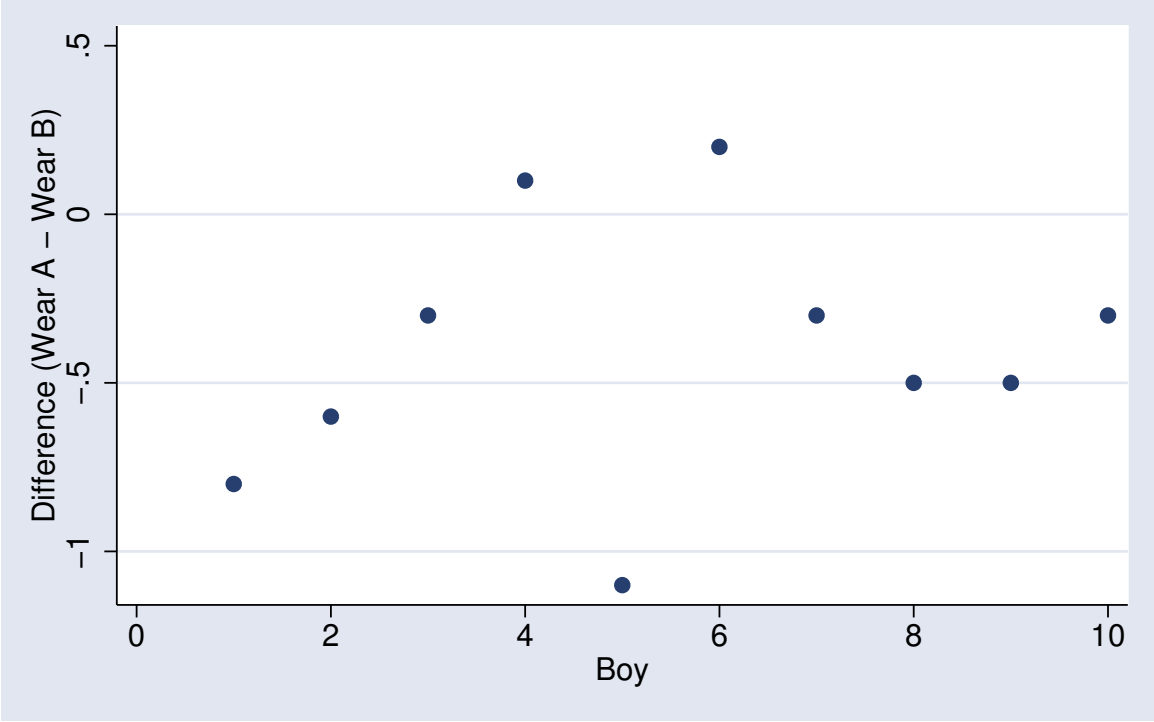
10 boys wearing special shoes. One shoe made with material A and the other with material B.



The approach to analysis in this design is to look at the difference in the responses for the two treatments.

Response: $x =$ Material A wear $-$ Material B wear

Then perform the appropriate one-sample t procedure on this difference.



```
. summarize diff

    Variable |       Obs        Mean    Std. Dev.       Min      Max
-------------+--------------------------------------------------------
        diff |        10   -.410000     .387155      -1.1       .2
```

A 95% CI for μ, the difference in mean wear is given by

$t^* = 2.262 \quad (df = 9)$

$$\begin{aligned} & -0.41 \pm 2.262 \frac{0.3872}{\sqrt{10}} \\ = \quad & -0.41 \pm 2.262 \times 0.1224 \\ = \quad & -0.41 \pm 0.2769 = (-0.687, -0.133) \end{aligned}$$

For testing $H_0 : \mu = 0$,

$$t = \frac{-0.41 - 0}{0.3872/\sqrt{10}} = \frac{-0.41}{0.1224} = -3.35$$

For $H_A : \mu \neq 0$ (as there is no reason to prefer one material),

$$p - \text{value} = 2 \times P[T \geq |-3.35|] = 2 \times 0.0043 = 0.0086$$

We have fairly good evidence that Material A tends to have less wear than Material B

```
. ttest diff == 0

One-sample t test

------------------------------------------------------------------------------
Variable |   Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    diff |    10  -.4100001    .1224292    .387155   -.6869541   -.1330461
------------------------------------------------------------------------------
Degrees of freedom: 9

                     Ho: mean(diff) = 0

     Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
       t =  -3.3489                 t =  -3.3489                  t =  -3.3489
   P < t =   0.0043             P > |t| =   0.0085              P > t =   0.9957
```

Assumptions underlying the paired t procedures

Need to differences in the observations to have a normal distribution for inference to be valid.

The actual observations do no need to be normally distributed.

Also for paired t tests, the null hypothesis is almost always $H_0 : \mu = 0$, though in theory it could be anything. Usually you are looking to see if there is a difference in the two treatments or groups and if it exists how big it is.