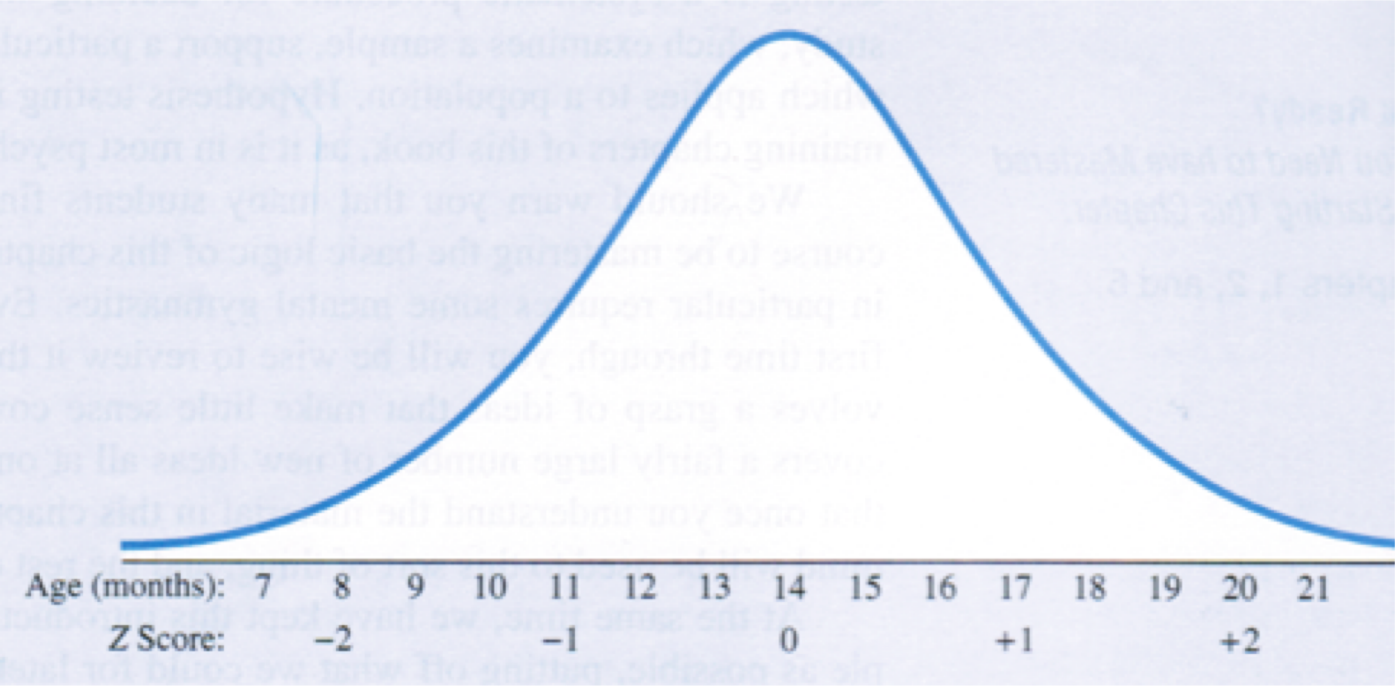
The Logic of Hypothesis Testing

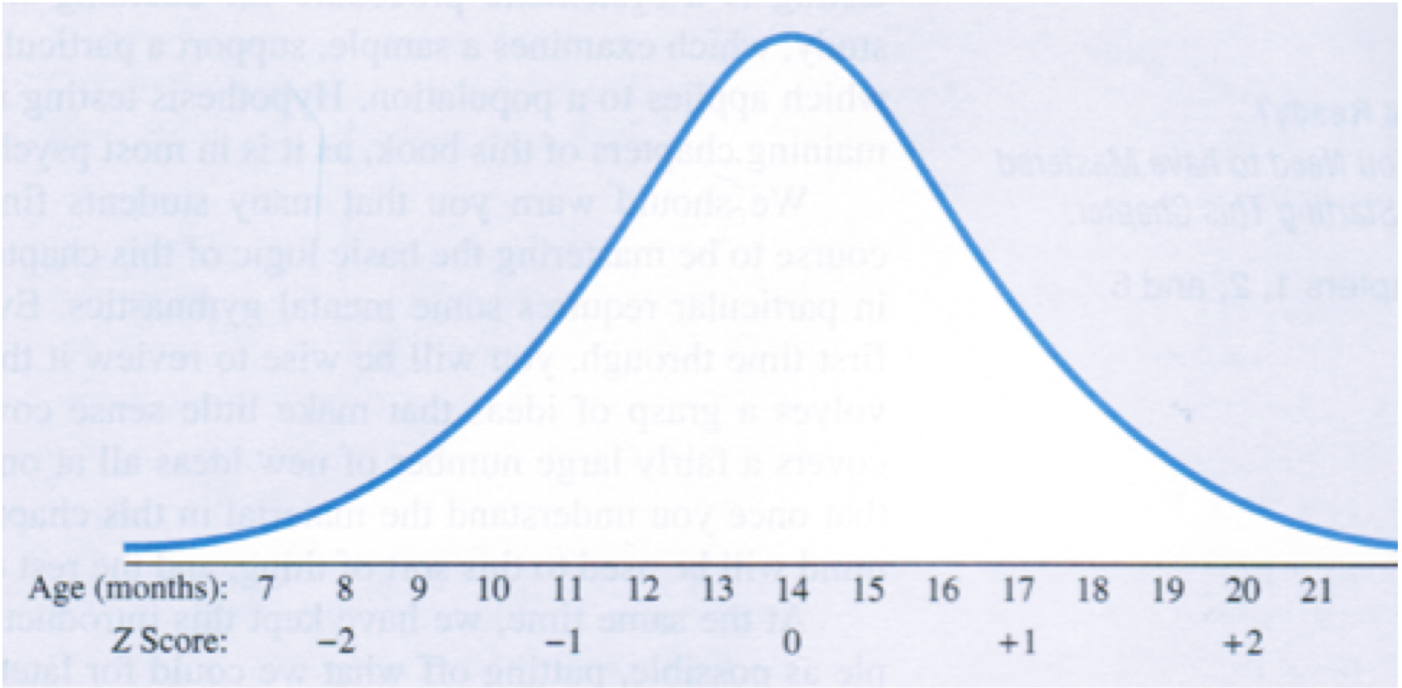
How do we know if our data differs from a population?

The Logic of Hypothesis Testing



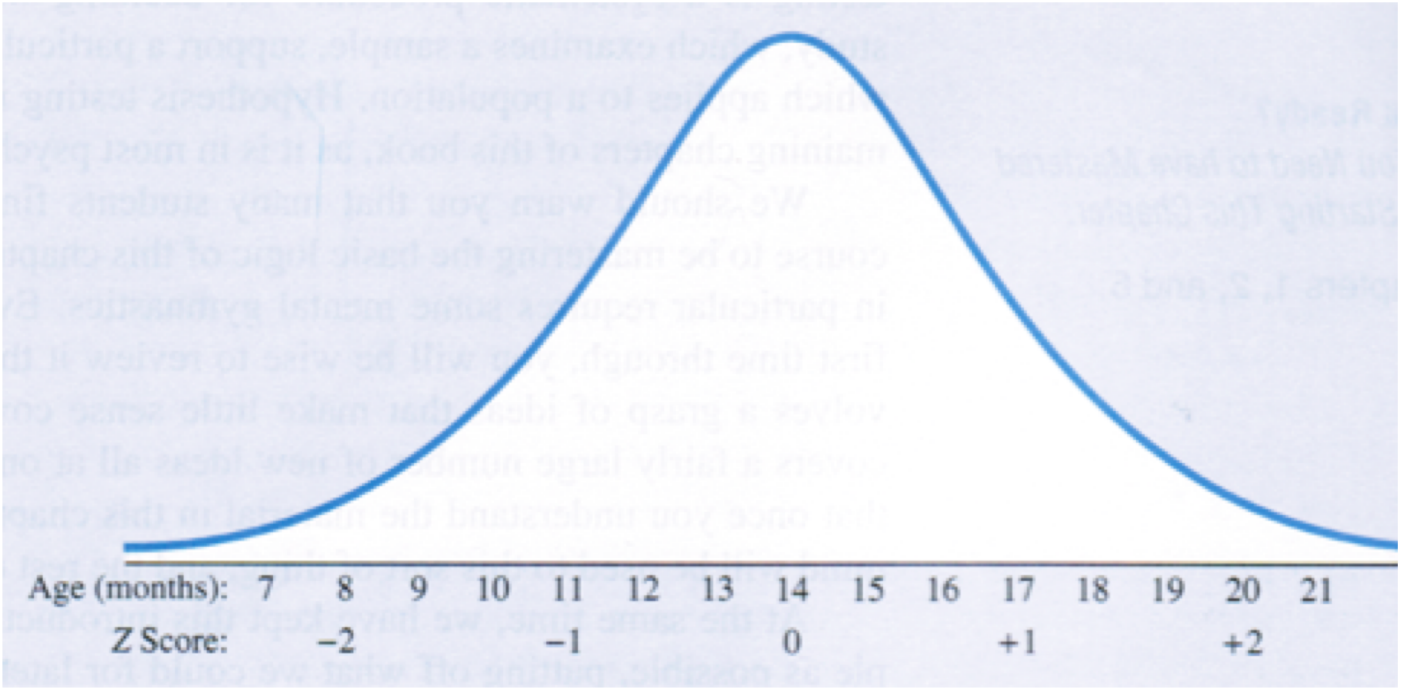
Let’s suppose the mean age for babies walking is 14 months.

We are going to try an experimental intervention that seeks to reduce that age.



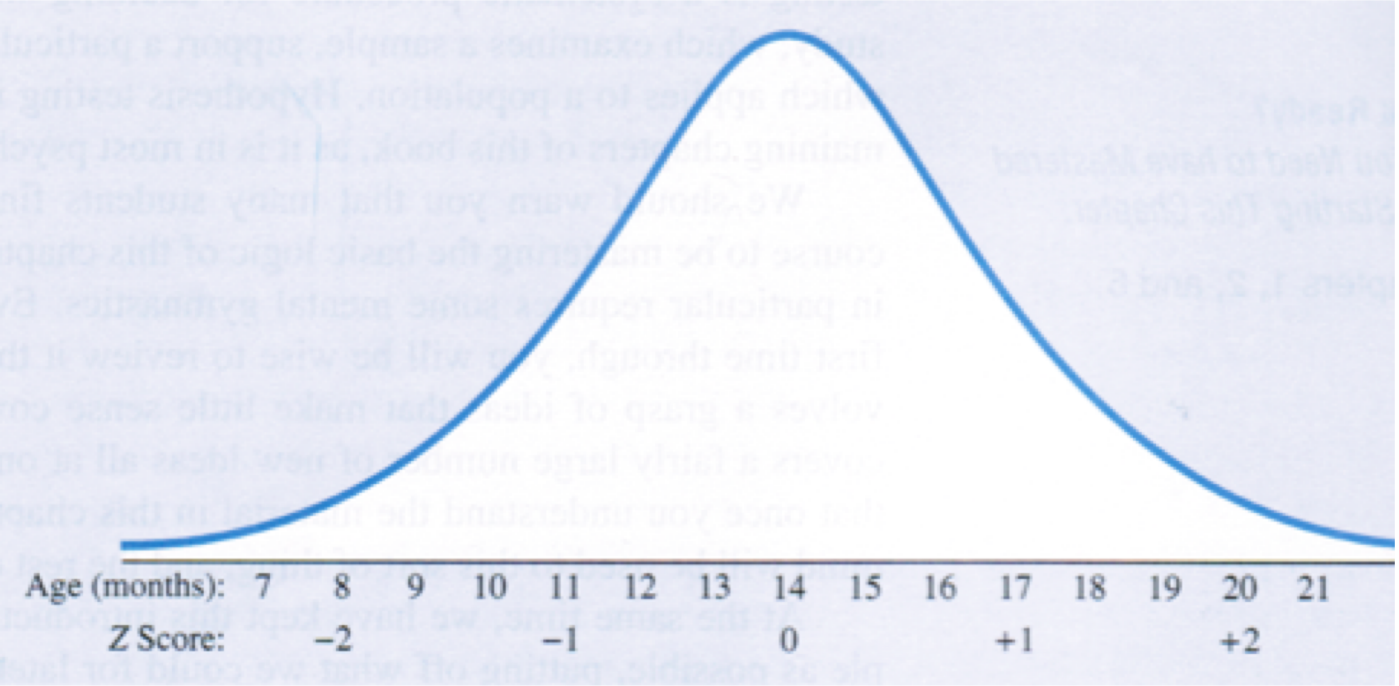
So we generate two hypotheses:

1) Our treatment does not work
2) Our treatment works



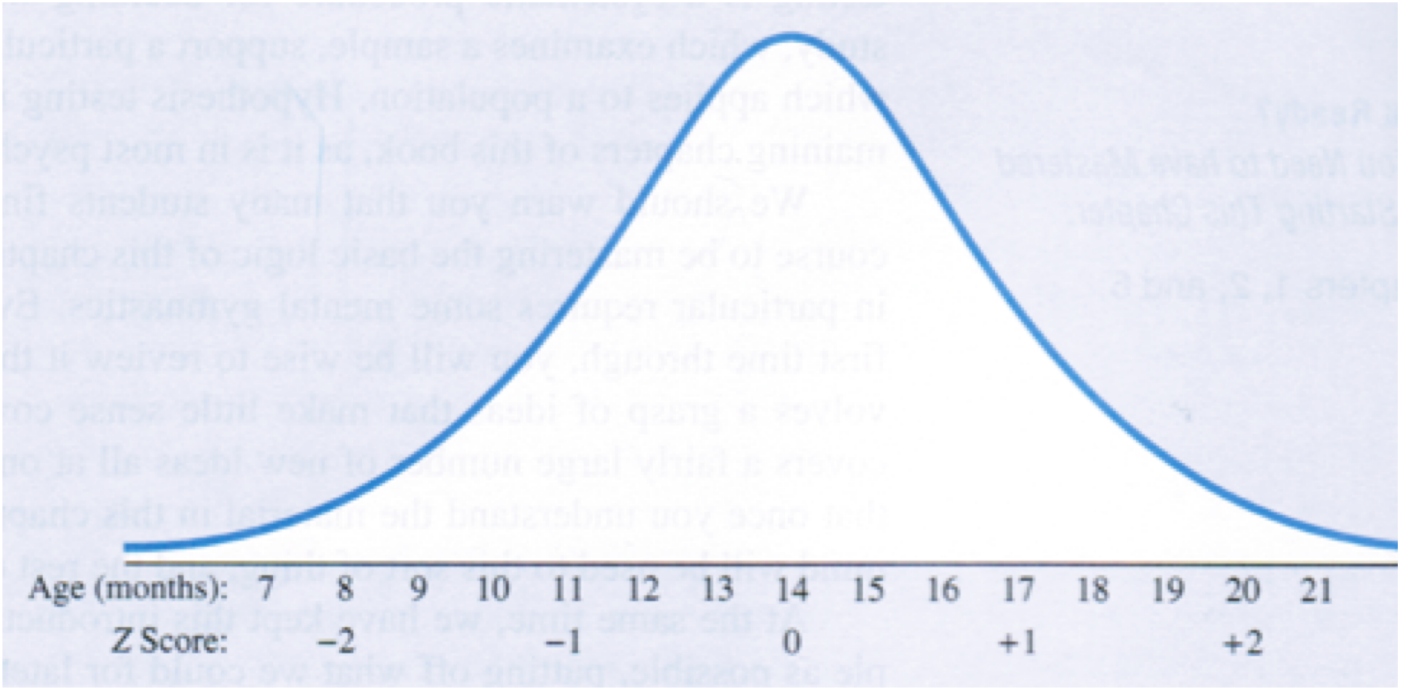
Lets think of this now in terms of populations:

Population One, babies who did not get our treatment
Population Two, babies who did get our treatment



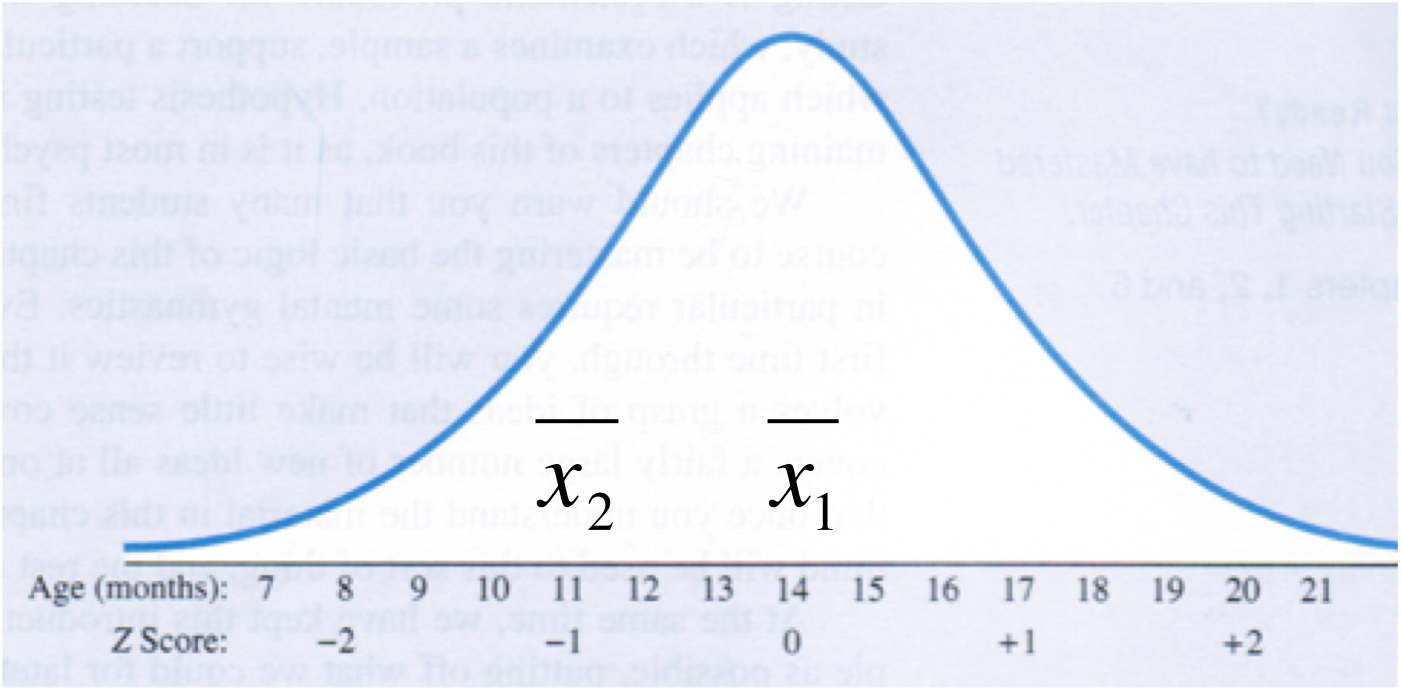
And we can quantify a descriptive statistic that is representative of our population, the mean age at which they begin walking

Population One: mean age of walking: μ_1
Population Two: mean age of walking: μ_2



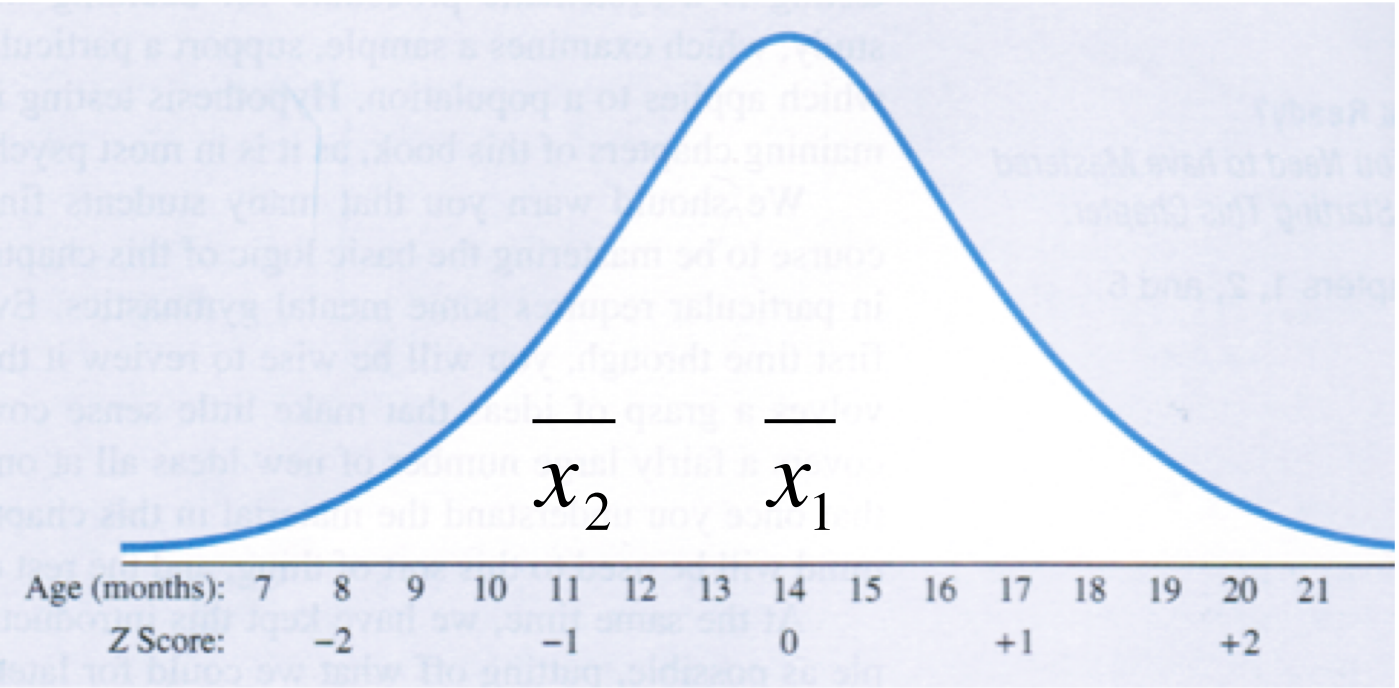
We typically frame this within the context of the null and alternative hypotheses:

1) H_0: The Null Hypothesis: Our treatment does not work: $\mu_1 = \mu_2$
2) H_1: The Alternative Hypothesis: Our treatment works: $\mu_1 \neq \mu_2$



So, let's say we run the study, and we obtain the following data.

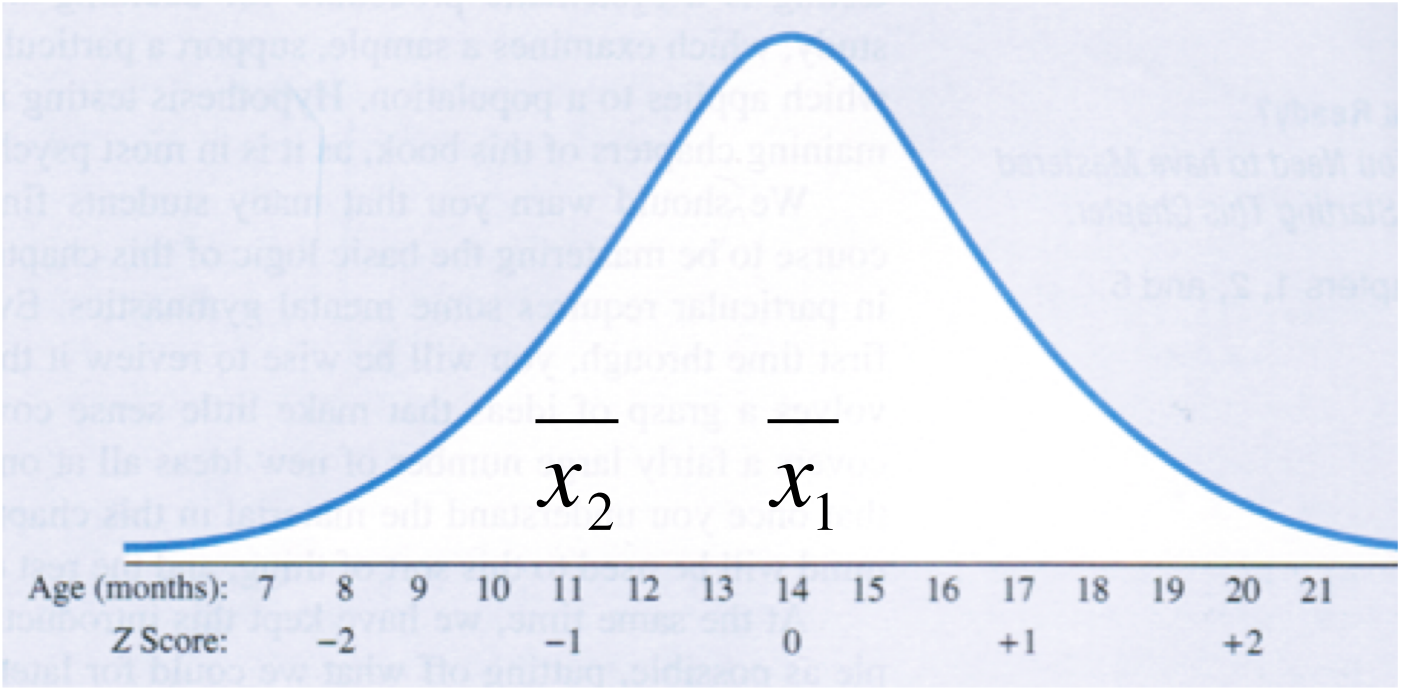
How do we know if our treatment worked?



Here's the weird bit, we first want to find out the probability of getting our result if the null hypothesis is true.

We call this our "critical value"

The probability of our results at which we will discard the null hypothesis and accept the alternative hypothesis.



0.05

The point at which a sample score is so extreme that we discard the null hypothesis

To do this we need to know the sampling distribution to perform statistical tests

The sampling distribution of the mean

provides all of the values the mean can take, along with the probability of getting each value if sampling is random from the null-hypothesis population

Consider the following…

Imagine you have a population that consists of 5 scores:

2, 3, 4, 5, 6

If you take samples of size 2, how many different possible samples are there and what do they consist of?

Sample		Sample	
1	2,2	14	4,5
2	2,3	15	4,6
3	2,4	16	5,2
4	2,5	17	5,3
5	2,6	18	5,4
6	3,2	19	5,5
7	3,3	20	5,6
8	3,4	21	6,2
9	3,5	22	6,3
10	3,6	23	6,4
11	4,2	24	6,5
12	4,3	25	6,6
13	4,4		

The sampling distribution of the mean

provides all of the values the mean can take, along with the probability of getting each value if sampling is random from the null-hypothesis population

Sample		$\bar{x}$	Sample		$\bar{x}$
1	2,2	2	14	4,5	4.5
2	2,3	2.5	15	4,6	5
3	2,4	3	16	5,2	3.5
4	2,5	3.5	17	5,3	4
5	2,6	4	18	5,4	4.5
6	3,2	2.5	19	5,5	5
7	3,3	3	20	5,6	5.5
8	3,4	3.5	21	6,2	4
9	3,5	4	22	6,3	4.5
10	3,6	4.5	23	6,4	5
11	4,2	3	24	6,5	5.5
12	4,3	3.5	25	6,6	6
13	4,4	4			

The sampling distribution of the mean

provides all of the values the mean can take, **along with the probability of getting each value if sampling is random from the null-hypothesis population**

Sample		$\bar{x}$	Sample		$\bar{x}$
1	2,2	2	14	4,5	4.5
2	2,3	2.5	15	4,6	5
3	2,4	3	16	5,2	3.5
4	2,5	3.5	17	5,3	4
5	2,6	4	18	5,4	4.5
6	3,2	2.5	19	5,5	5
7	3,3	3	20	5,6	5.5
8	3,4	3.5	21	6,2	4
9	3,5	4	22	6,3	4.5
10	3,6	4.5	23	6,4	5
11	4,2	3	24	6,5	5.5
12	4,3	3.5	25	6,6	6
13	4,4	4			

Probability of getting the mean

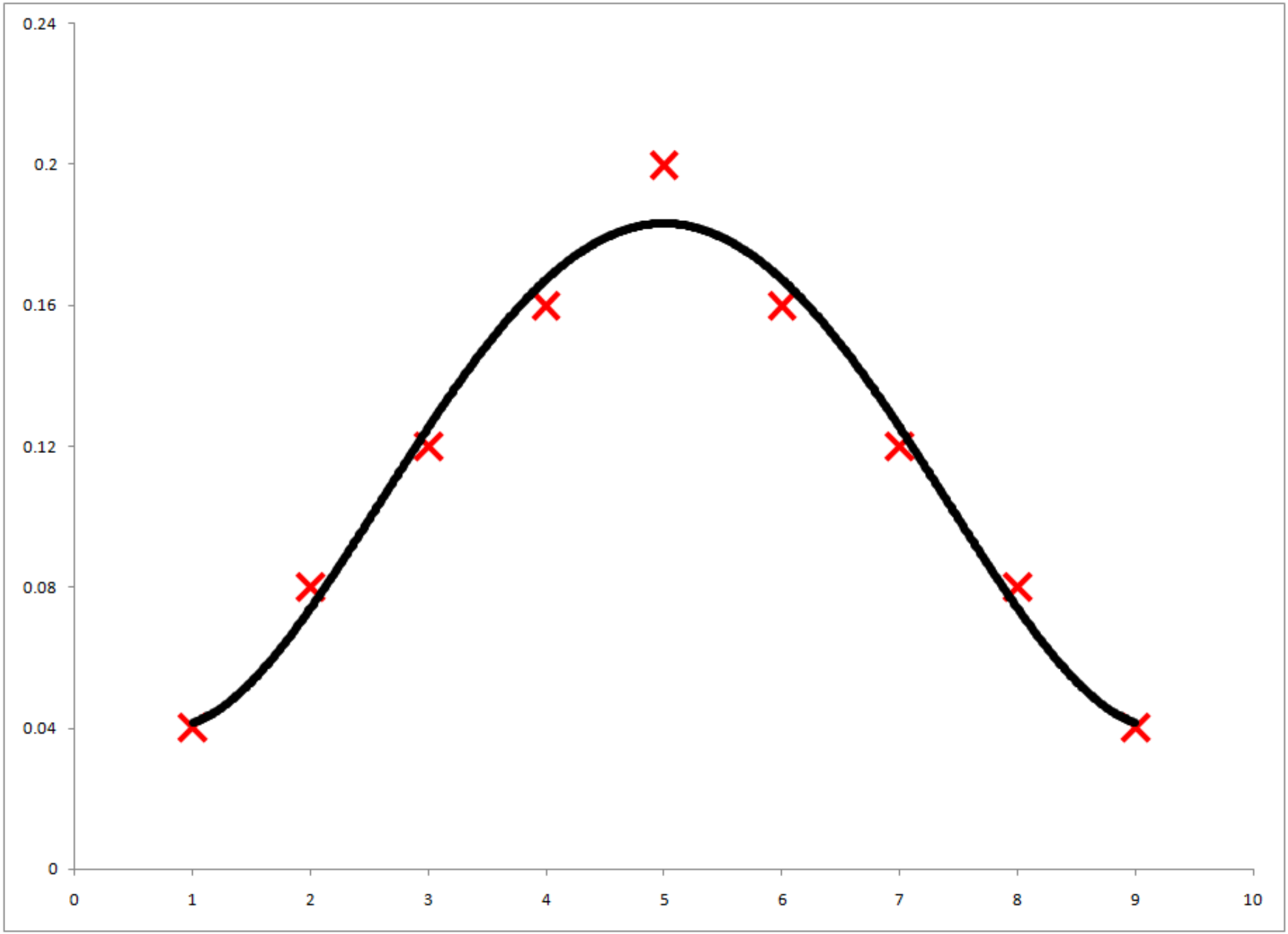
p ($\bar{x}$) = 2.0 1/25 0.04

Probability of getting the mean

$p(\bar{x}) = 2.0$	1/25	0.04
$p(\bar{x}) = 2.5$	2/25	0.08
$p(\bar{x}) = 3.0$	3/25	0.12
$p(\bar{x}) = 3.5$	4/25	0.16
$p(\bar{x}) = 4.0$	5/25	0.20
$p(\bar{x}) = 4.5$	4/25	0.16
$p(\bar{x}) = 5.0$	3/25	0.12
$p(\bar{x}) = 5.5$	2/25	0.08
$p(x) = 6.0$	1/25	0.04

Sampling distribution of the mean

$\bar{x}$	p($\bar{x}$)
2	0.04
2.5	0.08
3	0.12
3.5	0.16
4	0.20
4.5	0.16
5	0.12
5.5	0.08
6	0.04


0.24
0.2
0.16
0.12
0.08
0.04
0
0
1
2
3
4
5
6
7
8
9
10

Characteristics of the Sampling Distribution of the Mean

1. The sampling distribution of the mean has a mean ($\mu_{\bar{x}}$) and a standard deviation ($\sigma_{\bar{x}}$)

   Note than $\sigma_{\bar{x}}$ is also called the *standard error of the mean*

Characteristics of the Sampling Distribution of the Mean

2. The mean of the sampling distribution of the mean is the same as the mean of the population.

$$u_{\bar{x}} = u$$

Characteristics of the Sampling Distribution of the Mean

3. The standard deviation of the sampling distribution of the mean is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Characteristics of the Sampling Distribution of the Mean

4. The sampling distribution of the mean is normally shaped (usually)

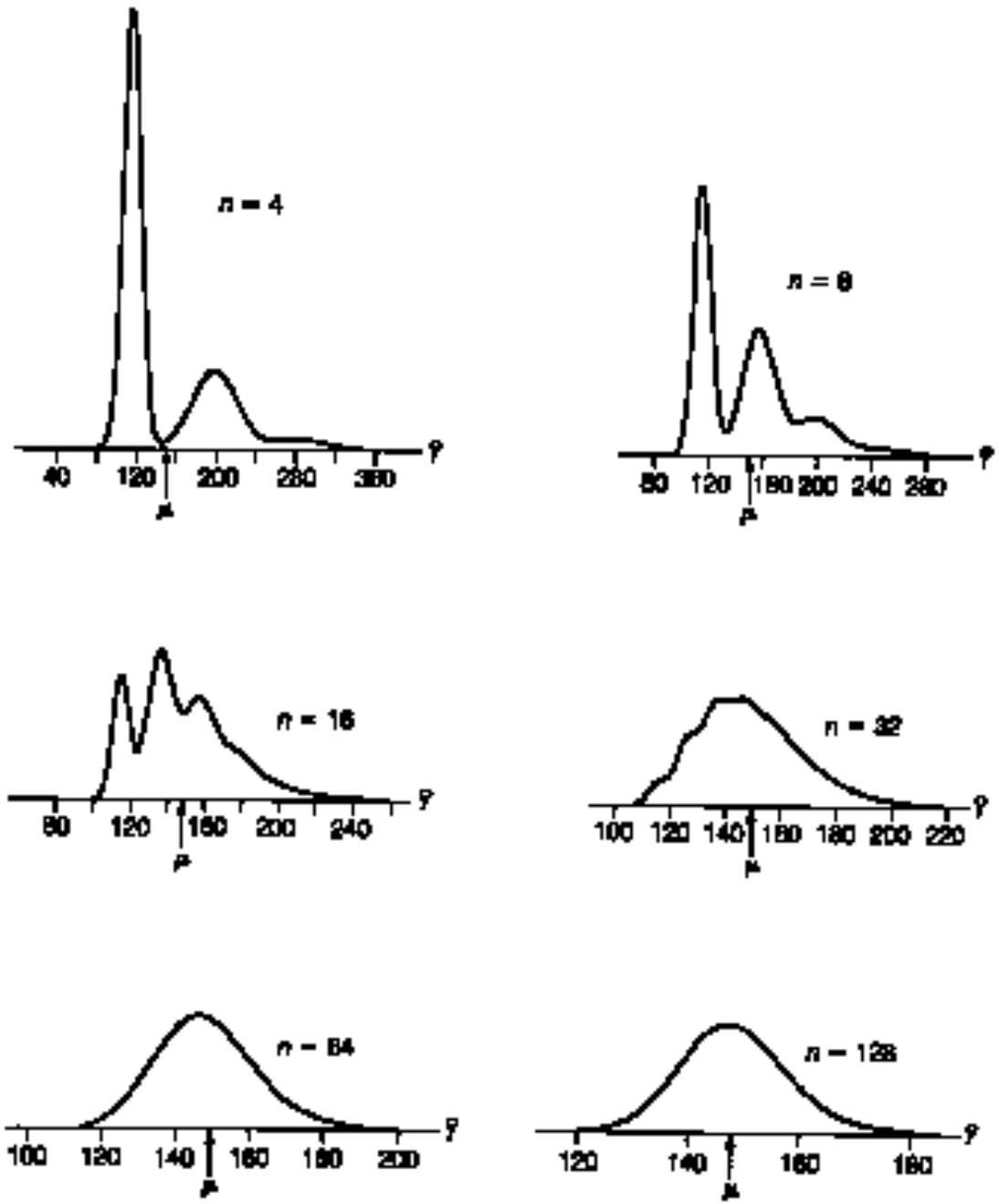
Characteristics of the Sampling Distribution of the Mean

4. The sampling distribution of the mean is normally shaped (usually)

The Central Limit Theorem

REGARDLESS of the shape of the population of raw scores, the sampling distribution of the mean approaches a normal distribution as sample size N increases

[redacted] 1.11
Sampling Distributions of $\bar{Y}$ for Samples from the Time-Score Population



The Binomial Distribution

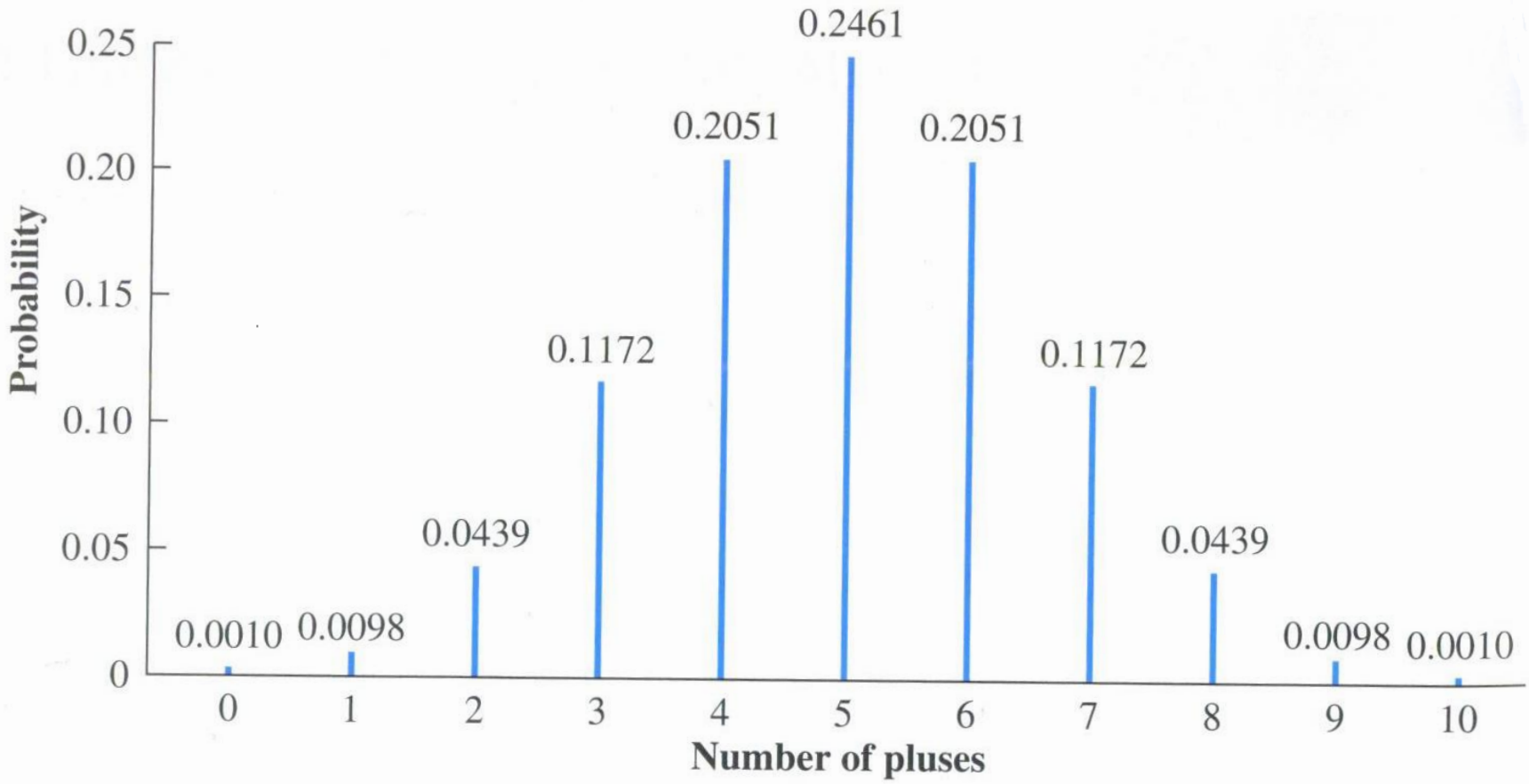
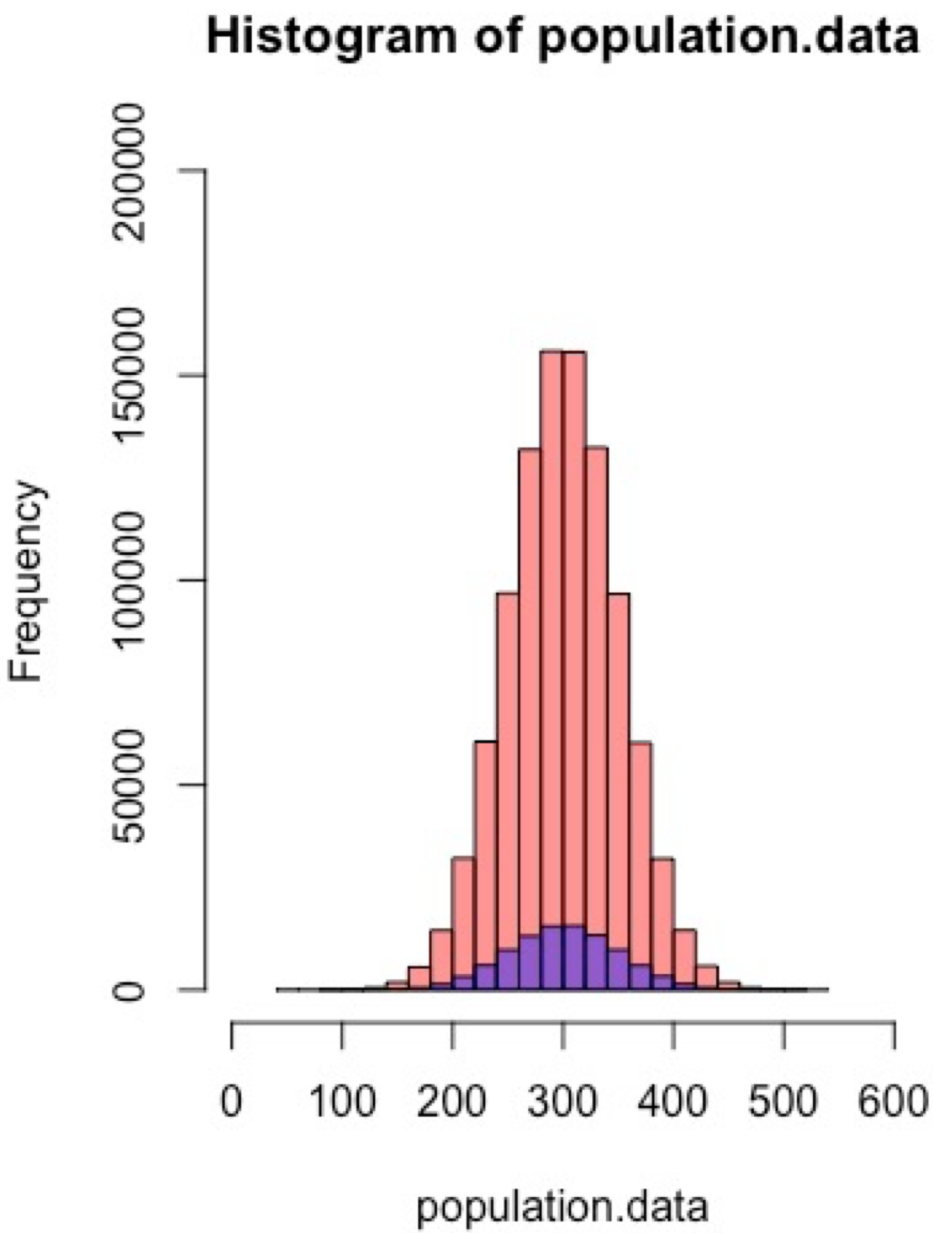


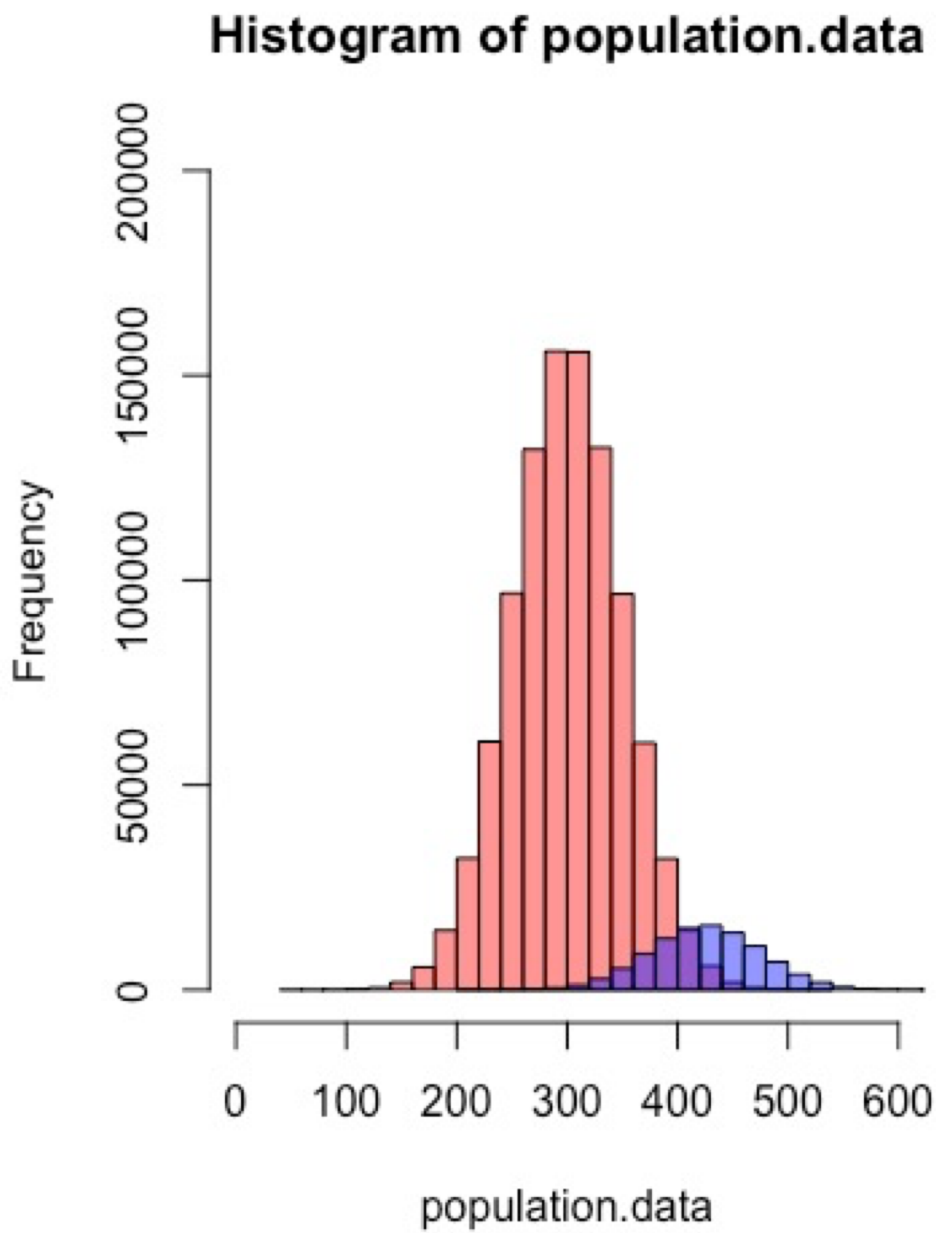
figure 10.1 Binomial distribution for $N = 10$ and $P = 0.50$.

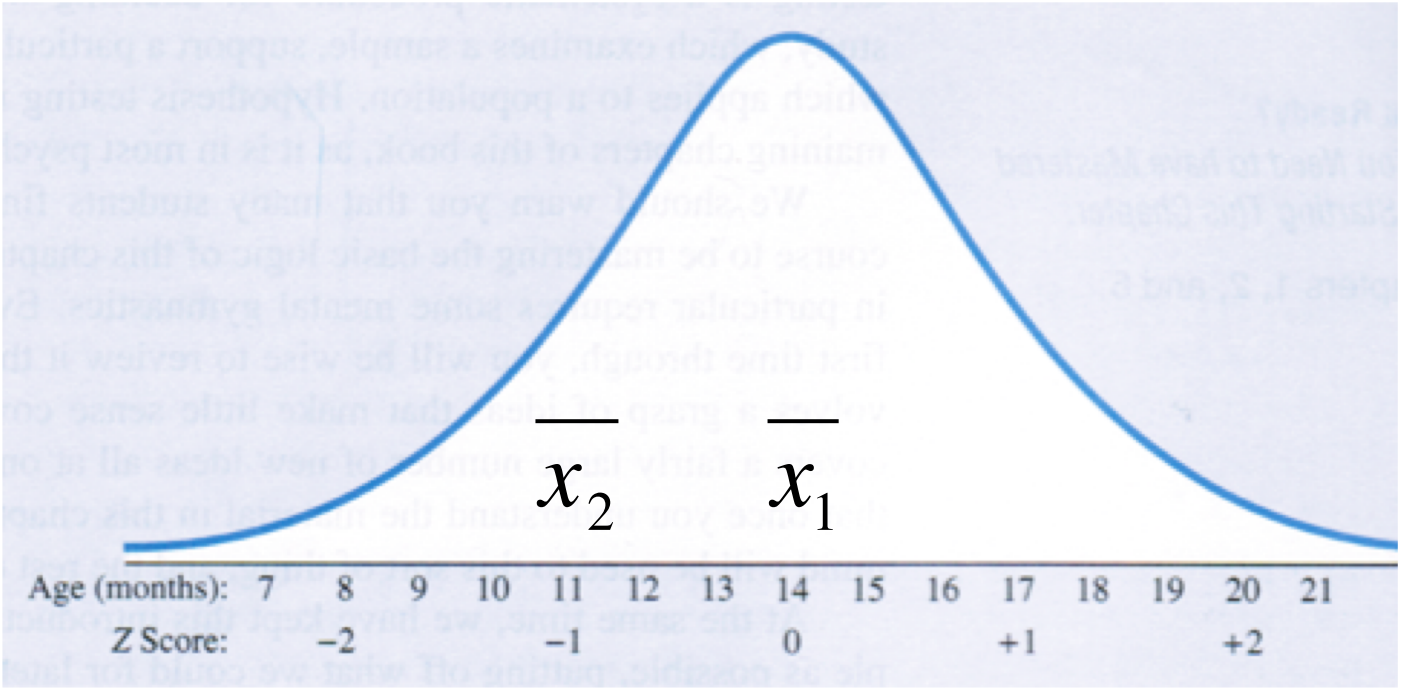
The Problem…

There is no way to know the sampling distribution of the mean for any data that we collect… so what do we compare our sample mean to???

What do we do?


Histogram of population.data
Frequency
0
50000
100000
150000
200000
0
100
200
300
400
500
600
population.data


Histogram of population.data
Frequency
0
50000
100000
150000
200000
0
100
200
300
400
500
600
population.data



We could just use the normal distribution…

WHY?

1. We know that the sampling distribution of the mean for any variable is normal given a reasonable sample size

2. We know the values and probabilities of the normal distribution

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

We could compare the data against a population with mean of 0 and standard deviation of 1 – the normal distribution… by doing this, we get the probability that the mean of our data comes from the normal distribution. But…

But…

The normal distribution is dead to use unless we know the standard deviation of the population…

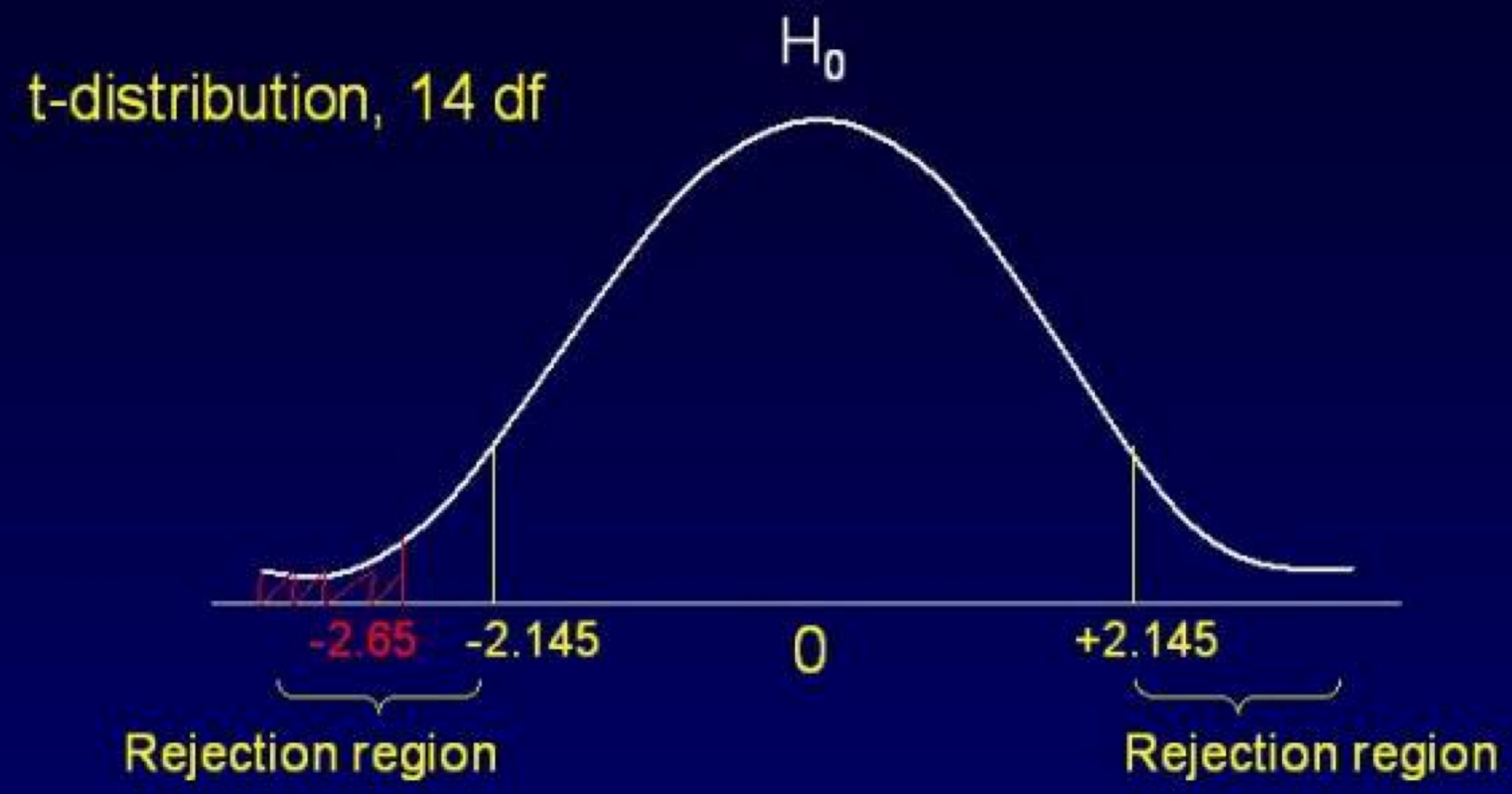
$$z = \frac{\text{sample mean} - \text{population mean}}{\text{population standard deviation}}$$

So what now?

What do we do?

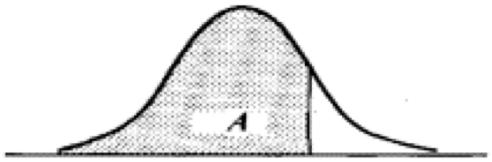
We compute some statistics <u>from our data</u> for which there is a known probability of scores given a specific sample size.

$$t = \frac{\text{sample mean} - \text{population mean}}{\text{standard deviation of sample}}$$



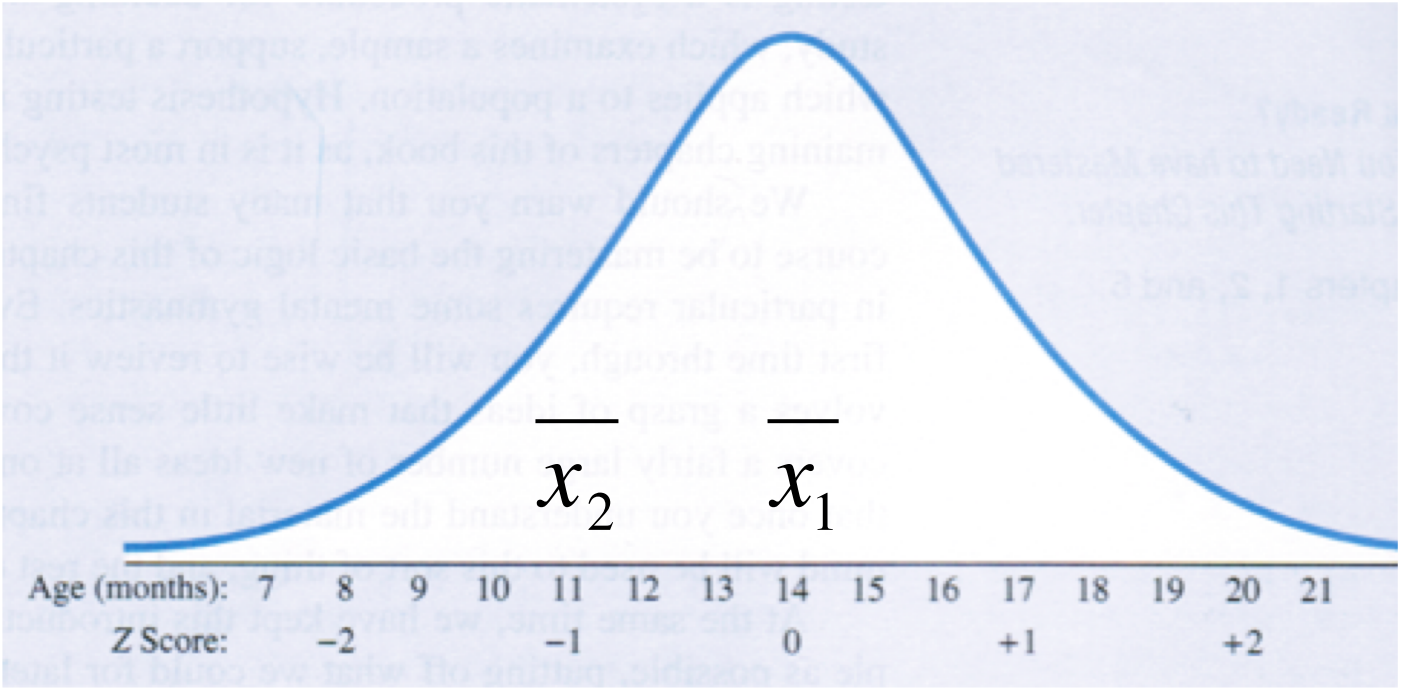
1) -2.65 falls in the rejection region – Reject the null
2) $0.01 < P < 0.02$ (2-sided test)

Entry is $t(A;\ \nu)$ where $P\{t(\nu) \le t(A;\ \nu)\} = A$



$t(A;\ \nu)$

	A						
ν	.60	.70	.80	.85	.90	.95	.975
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

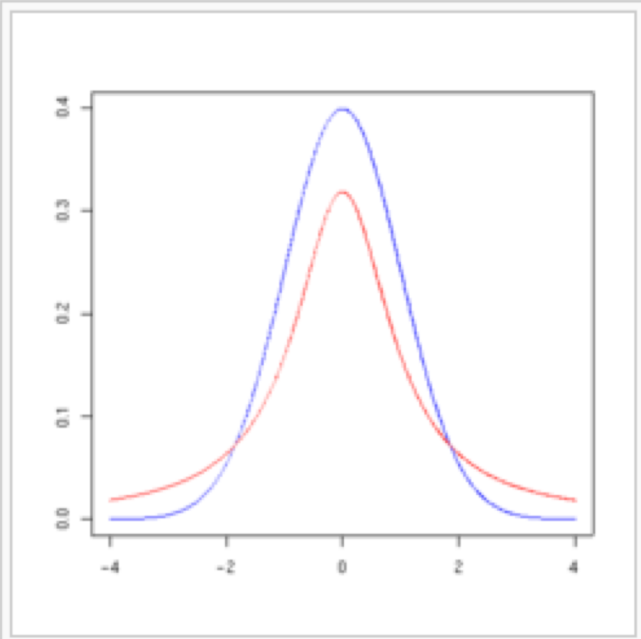


So what are we doing with z scores, t scores, F ratios, etc…

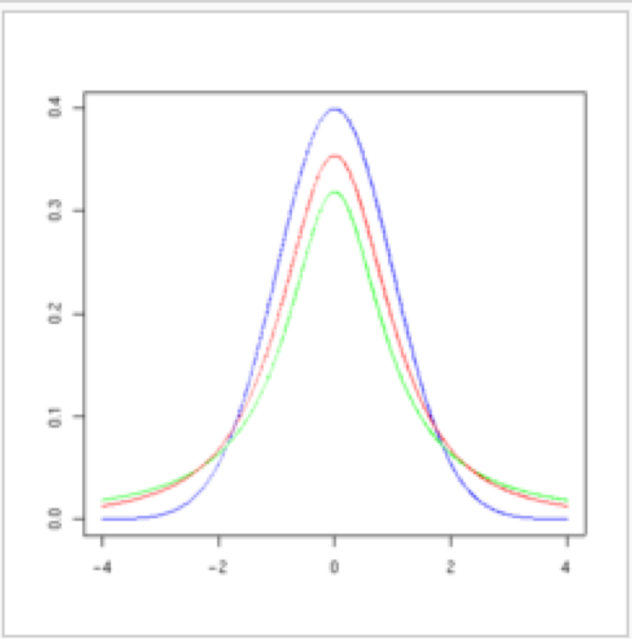
We are computing a statistic which is our samples score on the sampling distribution of the null hypothesis…

And WE ASSUME these magical distributions approximate the sampling distribution of the mean of our data…

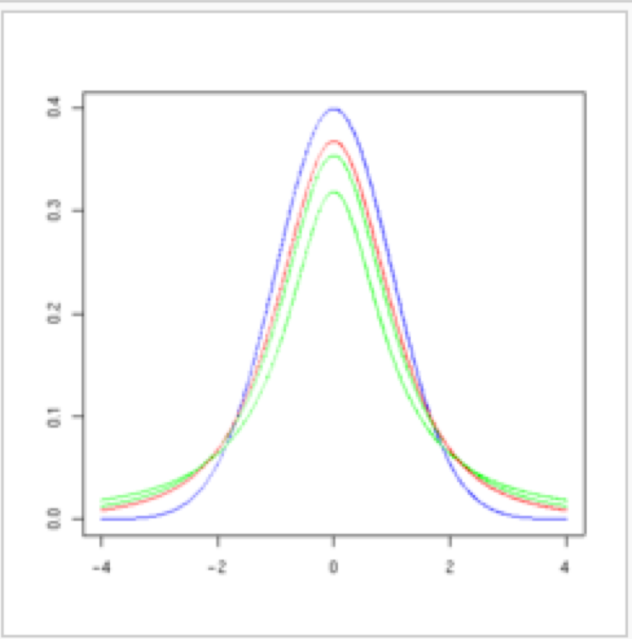
Density of the *t*-distribution (red) for 1, 2, 3, 5, 10, and 30 df compared to normal distribution (blue). Previous plots shown in green.



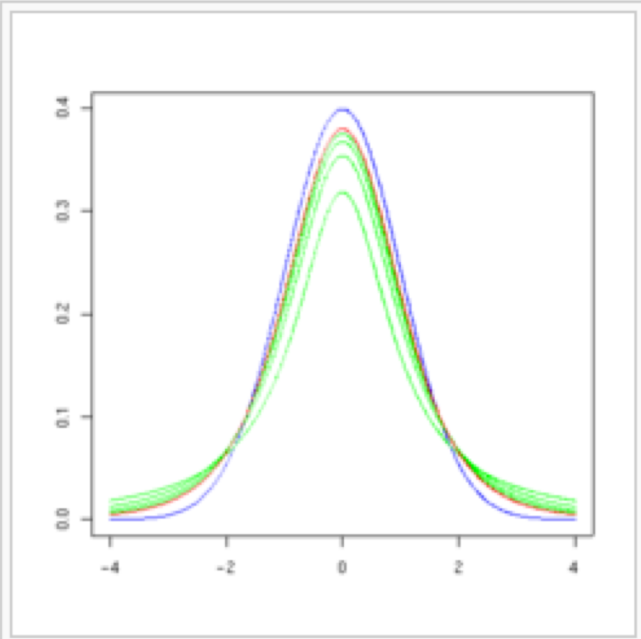
1 degree of freedom



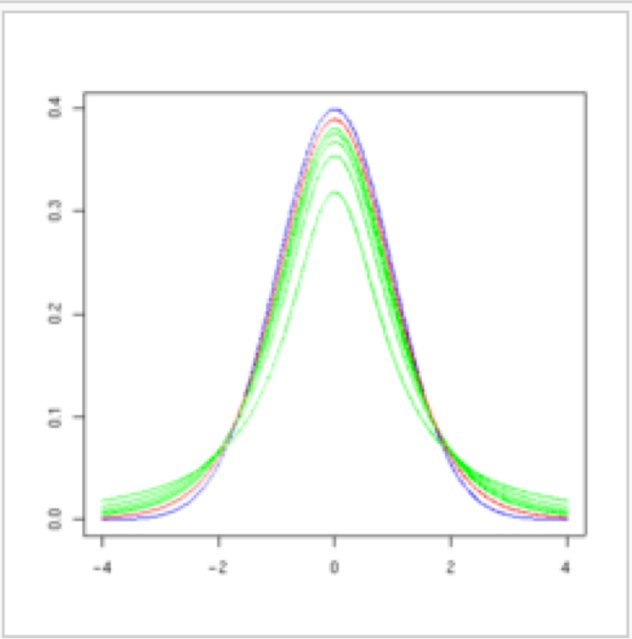
2 degrees of freedom



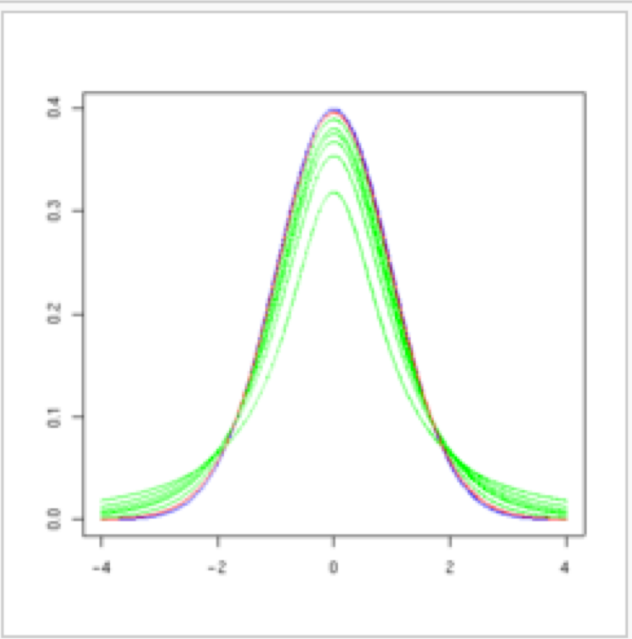
3 degrees of freedom



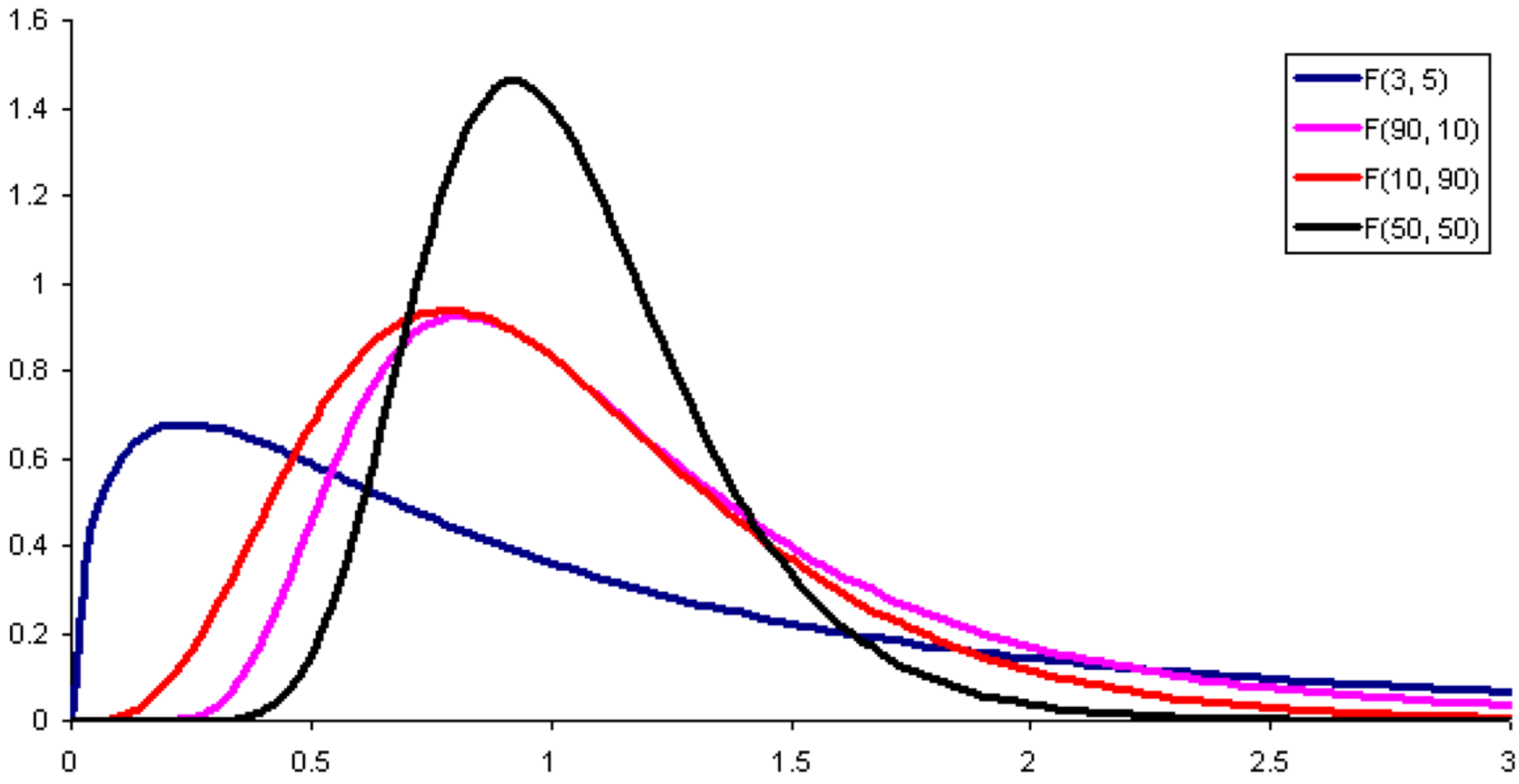
5 degrees of freedom



10 degrees of freedom



30 degrees of freedom


1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3
F(3, 5)
F(90, 10)
F(10, 90)
F(50, 50)

As stated, to draw these conclusions we need a test statistic to evaluation against the appropriate sampling distribution.

Test Statistic = variance explained by model

---

variance not explained by model

Test Statistic = effect

---

error

And then we can calculate the probability of getting the test statistic we have obtained and evaluating it against alpha.

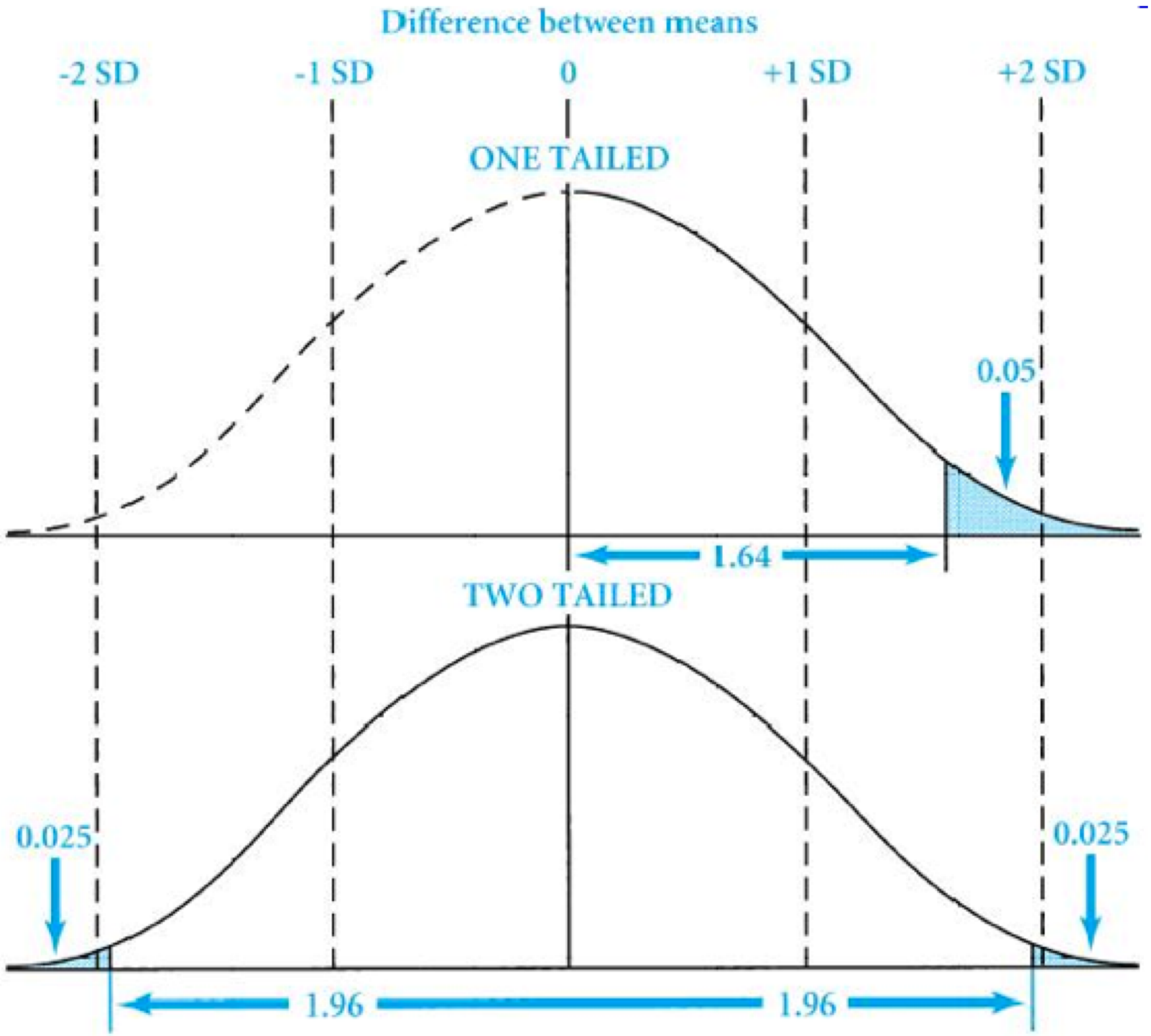
alpha = 0.05

if $p < 0.05$ then we say our test statistic is different

If $p > 0.05$ then we say our test statistic is the same

Other Things To Think About

One versus Two Tailed Significance


Difference between means
-2 SD
-1 SD
0
+1 SD
+2 SD
ONE TAILED
0.05
1.64
TWO TAILED
0.025
0.025
1.96
1.96

Type I and Type II Errors

<u>Type I Error</u>

Reject the null hypothesis when it is true.

<u>Type II Error</u>

Retain the null hypothesis when it is false.

State of Reality

Decision	Ho is true	Ho is false
Retain Ho	Correct Decision	Type II Error β
Reject Ho	Type I Error α	Correct Decision

Controlling for the possibility of Type I and Type II errors is not easy...

If we reduce alpha to reduce the chance of a Type I error, we increase the likelihood that we are making a Type II error!