Simple Statistical Tests

What are you testing?

- The General / Spouse / Cocktail Party level
  - I work on UAVs
  - I'm trying to get UAVs to fly even when GPS is not present
- The technical person level
  - I try to estimate pose (location and attitude) using a camera and possibly other sensors, but not GPS.
  - Difficult because:
    - Image processing takes a lot of computational power
    - Cameras are bearing only sensors
    - Cameras are very "noisy" sensors.
      - Images not the same with change of weather, etc.
      - Outliers
      - Algorithms used to obtain measurements

What are you testing?

- The advisor / colleague level
  - Image processing produces measurements with unknown uncertainty (covariance)
  - Need to generate estimates of covariance from image feature tracking data
- “Hypothesis” level
  - To estimate covariance, prior techniques used Kalman filters. Many robotics applications now use smoothing techniques.
    - H1. Can I estimate covariance using a smoothing technique
    - H2. Smoothing technique estimates will be more accurate than Kalman filter techniques

Partner work

- Pair up with someone else (preferably not someone you know very well).
- Walk through the four levels of “what you do” with each other

- One or two will randomly come up and present to the class

Presentations a week from Tuesday

- Find a well-respected paper:
  - ❑ Cited at least 20 times
  - ❑ Close to your research
- Analyze how the paper presents at the different levels (what is their hypothesis?)
- Present the 4 levels to the class

- In particular, pay attention to the related works section.
- You will be graded on how well other students understand you

Hypothesis testing

- What is a hypothesis test?

- You have a hypothesis:
  - Null-hypothesis starts with the assumption that you did nothing:

$$\mathcal{H}_0 : E[CovErr_K] = E[CovErr_S]$$

  - The alternate hypothesis is what you are trying to prove:

$$\mathcal{H}_1 : E[CovErr_K] > E[CovErr_S]$$

- If all samples are noisy, how do you do this?

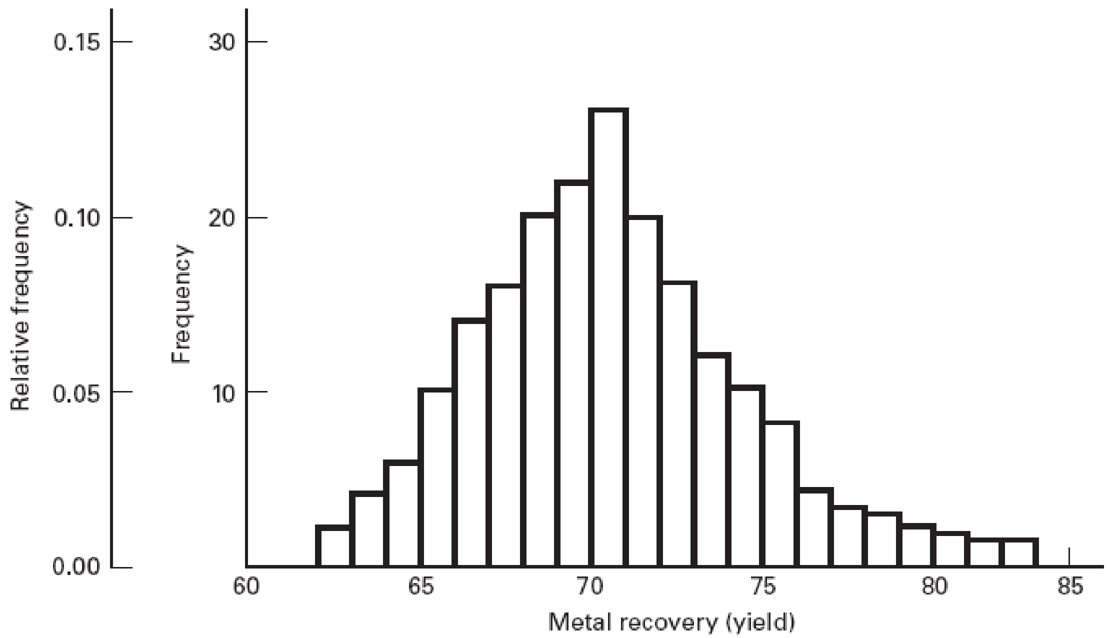
Errors:

- Type I error (false positive)
  - You say your method is better when it wasn't
- Type II error (false negative)
  - You say your method is not better when it was boxpl

- See MATLAB plots … is tst2(red) better than tst1(blue)
- Warm-up exercise
  - Generate two sets of random data, both from the same distribution
  - Assume you are trying to prove that set 2 is better (has a lower mean) than set 1.
  - Is set2 better than set 1?

2.2. Histogram

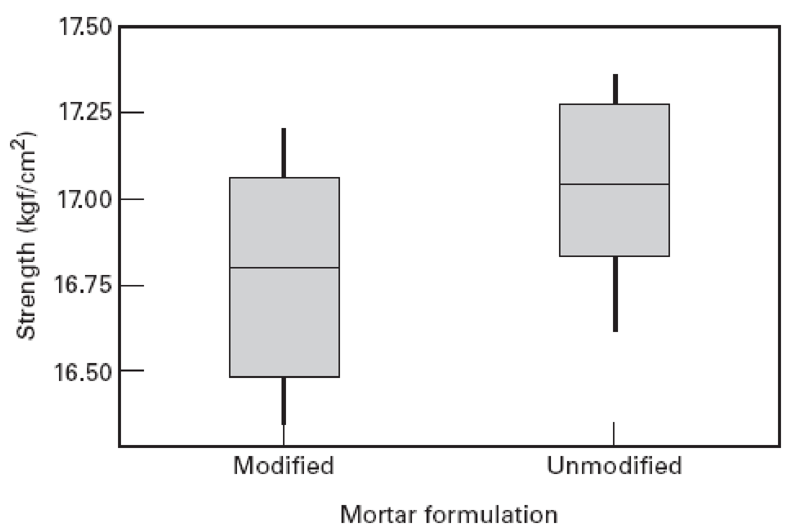
- useful for larger data sets
- Shows central tendency and the spread (variation) and the shape of the distribution



■ **FIGURE 2.2** Histogram for 200 observations on metal recovery (yield) from a smelting process

2.2. Box and Whisker Plot

- Displays maximum, minimum, lower and upper quartiles (25th and 75th percentiles) and median (50th percentile)
- Lines (whiskers) extends from the ends of the box to (typically) the minimum and maximum values



■ **FIGURE 2.3** Box plots for the Portland cement tension bond strength experiment

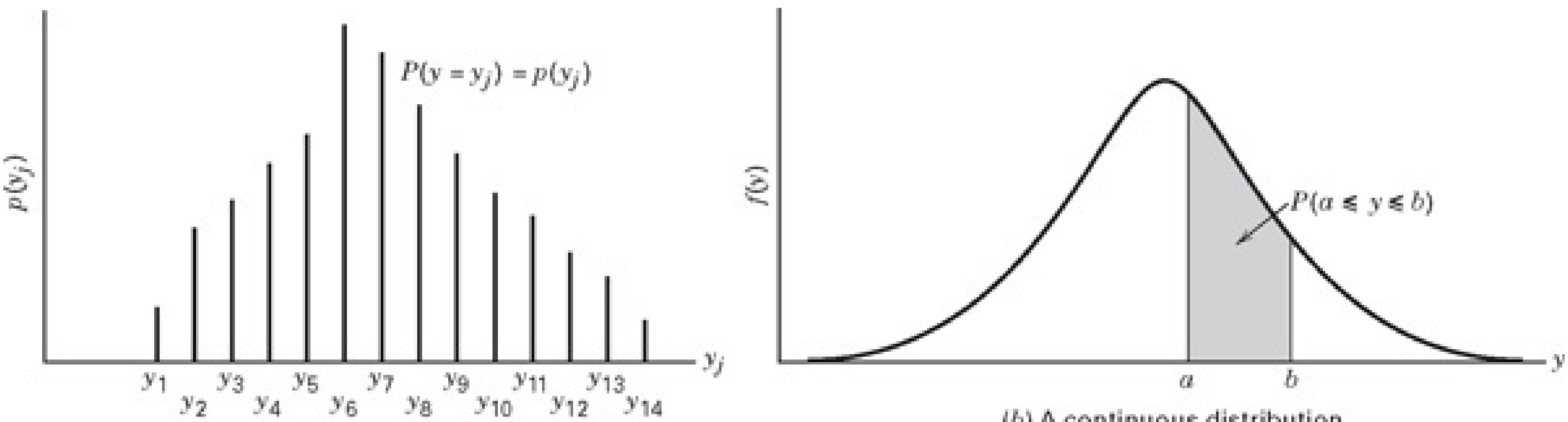
two samples of tension bond strength→ some difference in the mean strength. Both formulations have symmetric distributions with similar variability

2.2 Probability distribution

- Graphical tools useful to summarize info in a given sample of data.
- To describe observations that might occur in a sample more completely we use probability distributions
- A random variable y is described by its probability distribution

  discrete: $y \sim p(y_j)$ - probability mass function (pmf)

  continuous: $y \sim f(y)$ - probability density function (pdf)



(a) A discrete distribution

(b) A continuous distribution

Hypothesis testing

- Assume everything is Gaussian-like
  - Why? Central limit theorem
  - https://www.youtube.com/watch?v=Pujol1yC1_A&t=3s
- Choose a p-value
  - P-values are *not* a metric. They are a threshold and should only be used as such!
  - Probability such that the chance it came from the null hypothesis is very small

A simple test

- Assume
  - I want to test if the mean of some samples are 0.
  - I know the covariance of the values

- What is the p-value for 95% certainty?
- How do I test if the samples are zero-mean

- Run multiple tests. Ensure you get a Type 1 failure about 5% of the time. Also plot the p values for all runs

- Now let's remove the known covariance assumption

2.4.1.Hypothesis testing: One-Sample t

- Sample taken from population 1

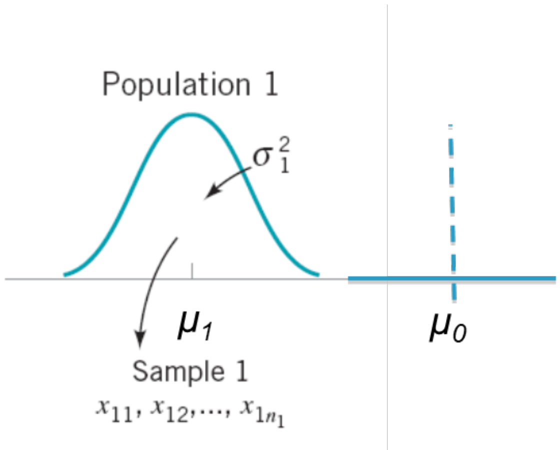
$$y_1, \ldots, y_n$$

- Sample statistics (mean and variance)

$$\bar{y}, \text{ and } s^2$$

- Assume: sample observations are Normal and Independently distributed

$$y_i \sim NID(\mu, \sigma^2)$$



- To test $H_0\colon \mu = \mu_0$ vs. $H_1\colon \mu \neq \mu_0$
- use test statistic

$$t_0 = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

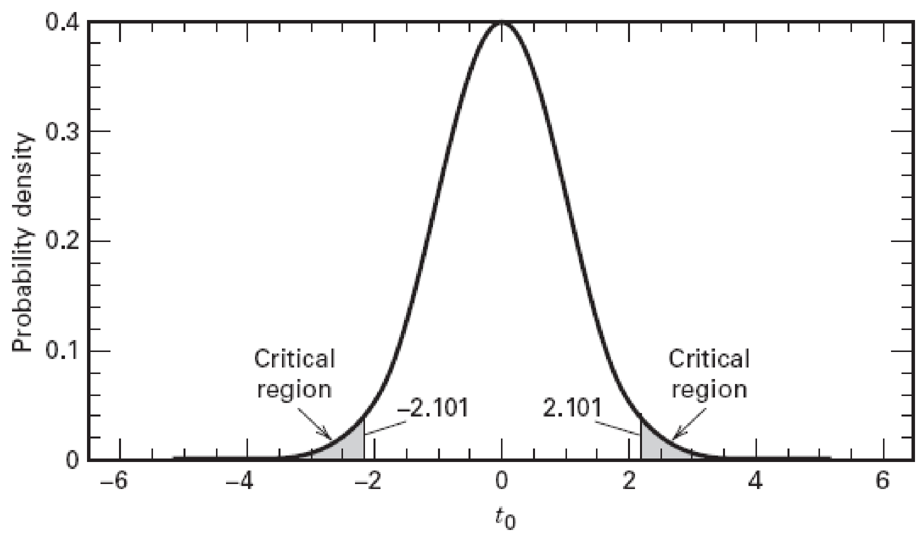
- Reject H_0 if $|t_0| > t(\alpha/2, n-1)$

2.4.1.Hypothesis testing: One-Sample t

- Under H_0, test statistic is distributed as $t_0 \sim t(n-1)$
- $100(1-\alpha)$ percent of t_0 fall between the rejection limits

$$-t_{\alpha/2,n-1} \text{ and } t_{\alpha/2,n-1}$$

- A t_0 outside rejection limits (in critical region) is unusual for H_0
  - H_0 rejected if $|t_0| > t_{\alpha/2,n-1}$ at α level of significance
  - Two sided alternative hypothesis



■ **FIGURE 2.10** The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$

- t distribution with 18 degrees of freedom
- For significance level $\alpha = 0.05$, percentage point is $t_{0.025,18} = 2.101$

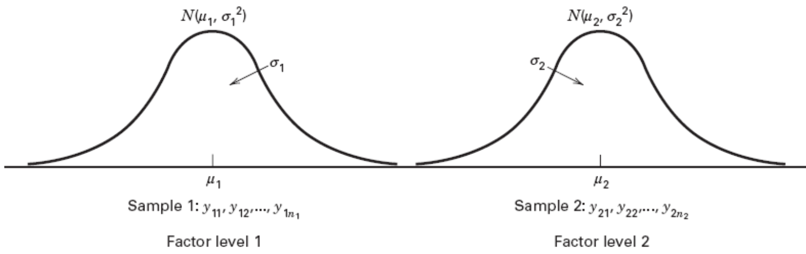
II Percentage Points of the t Distribution[a]

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = Degrees of freedom.

[a]Adapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

2.4.1. Hypothesis testing: Two-Sample t-Test



- Sample 1 (from population 1)

$$y_{11}, \cdots, y_{1,n1}$$

- Sample 2 (from population 2)

$$y_{21}, \cdots, y_{2,n2}$$

- Sample statistics (mean and variance)

$$\bar{y}_1, \bar{y}_2, s_1^2, s_2^2$$

- Assume: populations are Normal and Independently distributed with constant variance

$$y_{1j} \sim NID(\mu_1, \sigma^2)$$
$$y_{2j} \sim NID(\mu_2, \sigma^2)$$

- Test $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

2.4.1. Hypothesis testing: Two-Sample *t*-Test

- Use the test statistic

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- S_p^2 is the pooled sample variance (assumption $\sigma_1^2 = \sigma_2^2$)

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

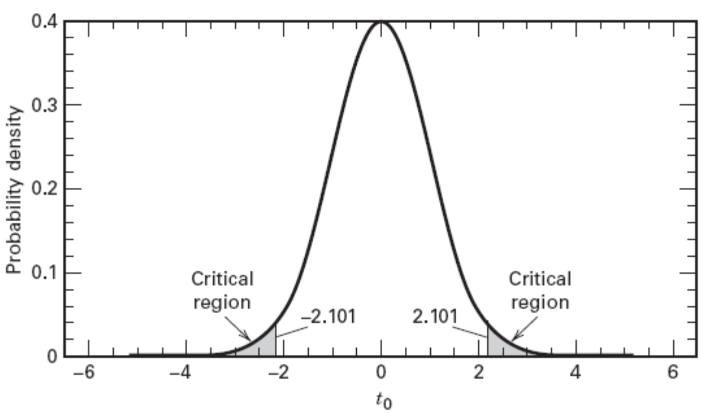
- Under H_0, the test statistic is distributed as

$$t_0 \sim t(n_1 + n_2 - 2)$$

  - How far apart the averages are, expressed in standard deviation units
  - Values of t_0 near 0 consistent with $H_0: \mu_1 = \mu_2$

One-sided alternative hypothesis

- H_0 rejected if $|t_0| > t_{\alpha/2,n_1+n_2-2}$ at α level of significance
- If we want to test H_1: $\mu_1 > \mu_2$
  - Reject H_0 only if $t_0 > t_{\alpha,n_1+n_2-2}$
- If we want to test H_1: $\mu_1 < \mu_2$
  - Reject H_0 only if $t_0 < -t_{\alpha,n_1+n_2-2}$



■ **FIGURE 2.10** The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$

For H_1: $\mu_1 \neq \mu_2$
Use $t_{0.025,18} = 2.101$
Reject H_0 if $|t_0| > 2.101$

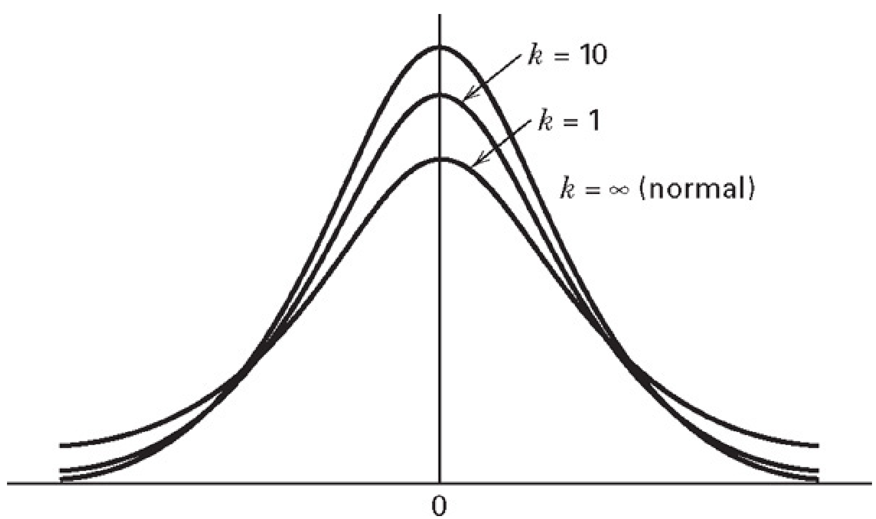
For H_1: $\mu_1 > \mu_2$
Use $t_{0.05,18} = 1.734$
Reject H_0 if $t_0 > 1.734$

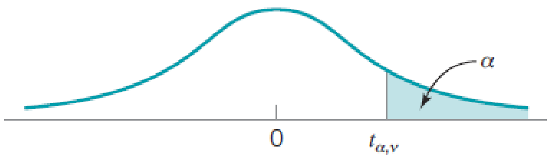
2.3. Sampling Distributions-Student's t

- Student's t probability density function
  - y is a Student's t random variable with k degrees of freedom

$$y \sim t_k \text{ or } t(k)$$

- Table II in Appendix →CDF of t distribution





II Percentage Points of the t Distribution[a]

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = Degrees of freedom.

[a]Adapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

2.3. Sampling Distributions-Student's t

- If $z \sim N(0,1)$ and $x \sim \chi^2(k)$, then

$$t = \frac{z}{\sqrt{x/k}} \sim t(k)$$

- As an example, consider sample mean $\bar{y}$ and sample variance s^2 of a random sample from $y_i \sim NID(\mu, \sigma^2)$
- Then the test statistic t_0 has the sampling distribution

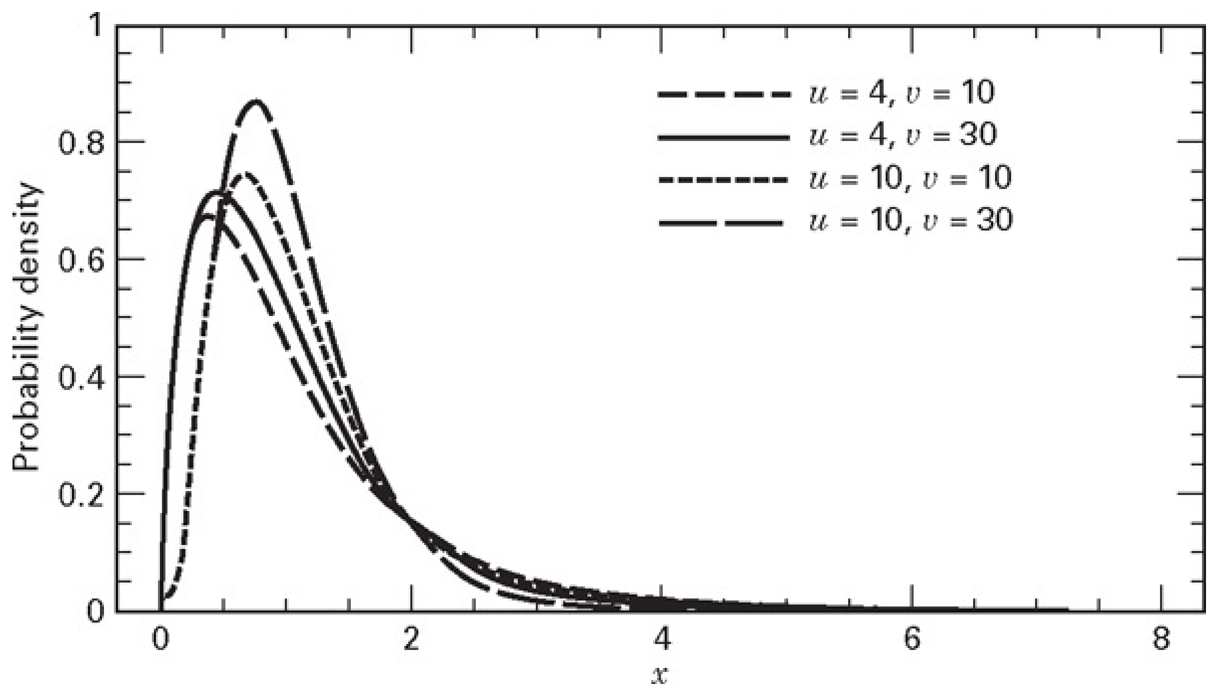
$$t_0 = \frac{\bar{y} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

2.3. Sampling Distributions- F

- F probability density function
  - ❑ y is an F random variable with u numerator d.f. and v denominator d.f.

$$y \sim F(u, v) \text{ or } F_{u,v}$$

- Table IV in Appendix →CDF of F distribution



2.3. Sampling Distributions- F

- If $x_1 \sim \chi^2(u)$ and $x_2 \sim \chi^2(v)$

$$y = \frac{x_1/u}{x_2/v} \sim F(u, v)$$

- Suppose $y_1 \sim N(\mu_1, \sigma^2)$ and $y_2 \sim N(\mu_2, \sigma^2)$ are two populations with common variance. Random samples taken as

$$y_{11}, y_{12}, \dots, y_{1n_1} \text{ and } y_{21}, y_{22}, \dots, y_{2n_2}$$

- Let s_1^2 and s_2^2 be sample variances. The test statistic s_1^2/s_2^2 has the sampling distribution

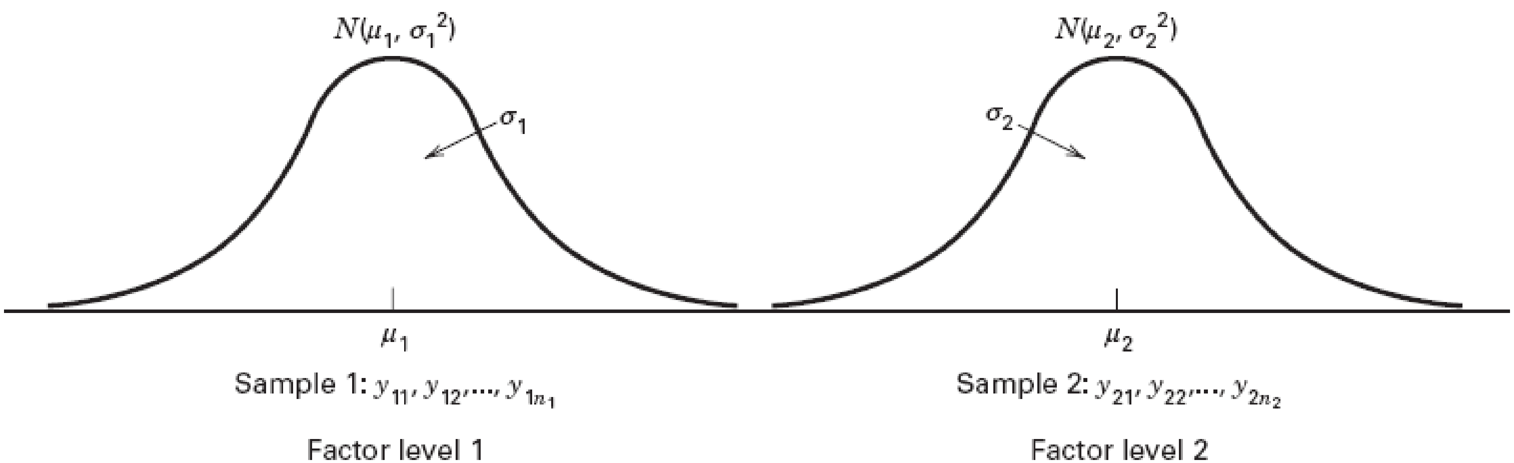
$$\frac{s_1^2}{s_2^2} = \frac{\left(\frac{SS_1}{\sigma^2}\right)/(n_1-1)}{\left(\frac{SS_2}{\sigma^2}\right)/(n_2-1)} \sim F(n_1-1, n_2-1)$$

Hypothesis testing … reminder

- You can only disprove the null hypothesis (with some probability. You can't:
  - ❑ Prove the null hypothesis
  - ❑ Use the same data repeatedly to find what is true,
    - o https://xkcd.com/882/

2.4. Inferences about Differences in Means

- Simple comparative experiment: Make inference about the means of two populations or treatments: two sample t-test
  - ❑Based on random sample from the two populations
  - ❑Inference: hypothesis testing and confidence interval

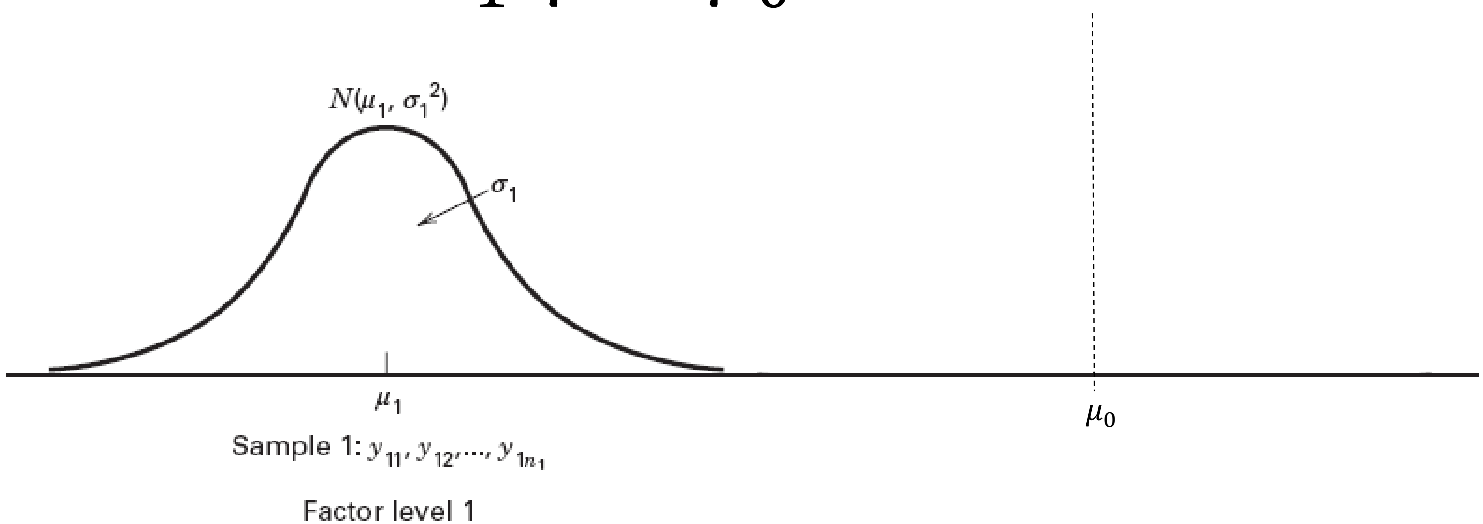


■ **FIGURE 2.9** The sampling situation for the two-sample *t*-test

2.4. Inferences about Differences in Means

- Inference about the mean of one population: determine whether mean μ of a population is equal to a specified value μ_0
  - One-sample t- test. Hypotheses to be tested:

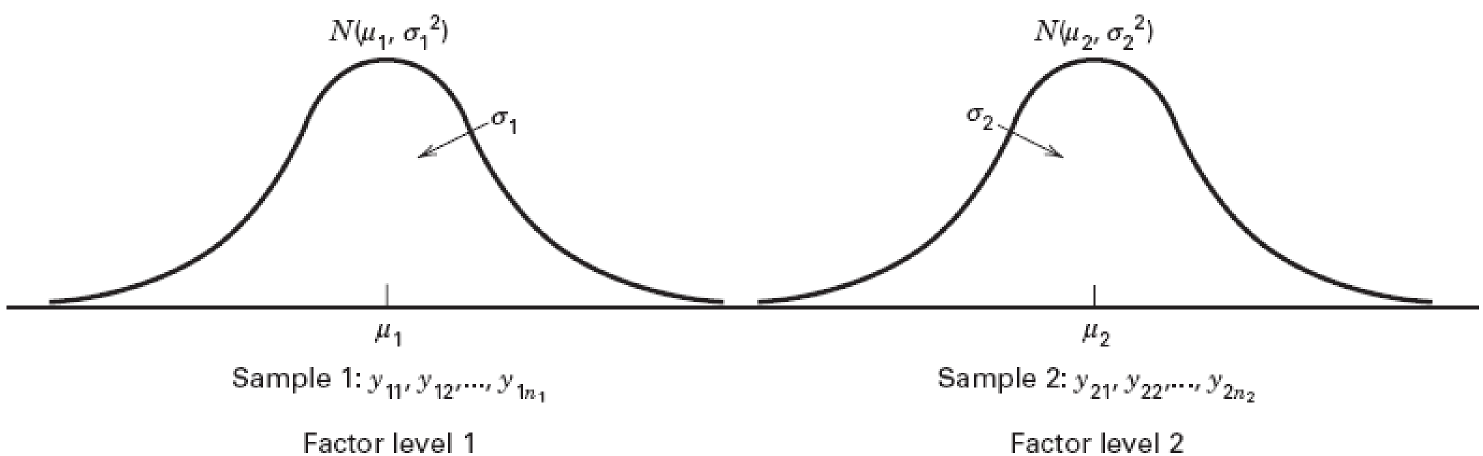
$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$



2.4. Inferences about Differences in Means

- Simple comparative experiment: determine whether the means μ_1 and μ_2 of two populations are different
  - Two-sample t-test. Hypotheses to be tested:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$



2.4.1. Hypothesis testing

- Take a random sample of observations from the populations
- Compute a test statistic that measures the difference in means
- Decision rule
  - If observed test statistic falls into critical (or rejection) region then <u>we reject H_0</u> → observed difference is statistically significant; the treatments are different
  - If observed test statistic does not fall into critical region <u>we fail to reject H_0</u> → observed difference is due to chance; the treatments are not different

2.4.1. Hypothesis testing

- Two kinds of errors may be committed
  - ❑ H_0 is rejected when it is true → a type I error has occurred

$$\alpha = Pr(\text{type I error})$$

  - ❑ H_0 is not rejected when false → a type II error has occurred

$$\beta = Pr(\text{type II error})$$

- Specify α → determines the critical region
  - ❑ α: level of significance of the test
- Choose sample size (n_1, n_2) so that β is suitably small

2.4.1.Hypothesis testing: One-Sample t

- Sample taken from population 1

$$y_1, \dots, y_n$$

- Sample statistics (mean and variance)

$$\bar{y}, \text{ and } s^2$$

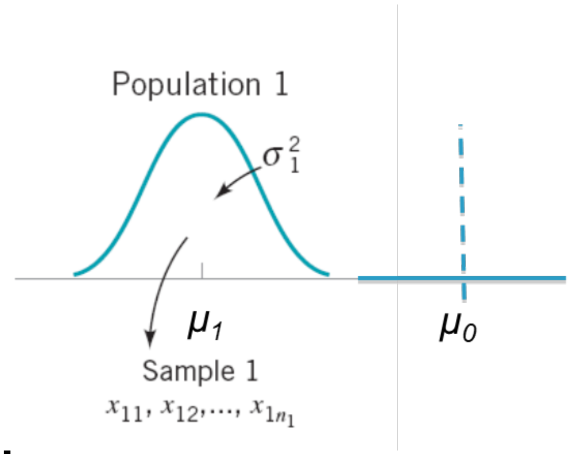
- Assume: sample observations are Normal and Independently distributed

$$y_i \sim NID(\mu, \sigma^2)$$

- To test $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$
- use test statistic

$$t_0 = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

- Reject H_0 if $|t_0| > t(\alpha/2, n-1)$

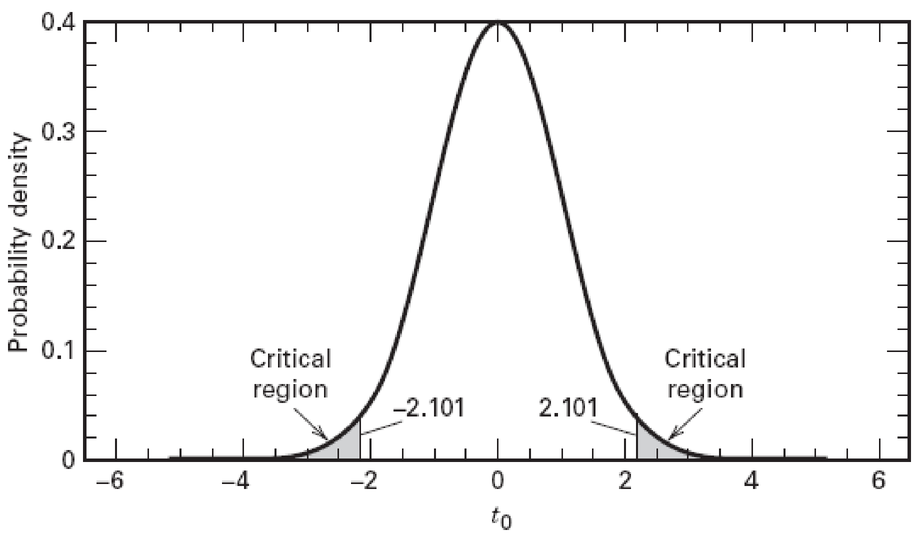


2.4.1.Hypothesis testing: One-Sample t

- Under H_0, test statistic is distributed as $t_0 \sim t(n-1)$
- $100(1-\alpha)$ percent of t_0 fall between the rejection limits

$$-t_{\alpha/2,n-1} \text{ and } t_{\alpha/2,n-1}$$

- A t_0 outside rejection limits (in critical region) is unusual for H_0
  - H_0 rejected if $|t_0| > t_{\alpha/2,n-1}$ at α level of significance
  - Two sided alternative hypothesis



■ **FIGURE 2.10** The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$

- t distribution with 18 degrees of freedom
- For significance level $\alpha = 0.05$, percentage point is $t_{0.025,18} = 2.101$

II Percentage Points of the t Distribution[a]

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = Degrees of freedom.

[a]Adapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

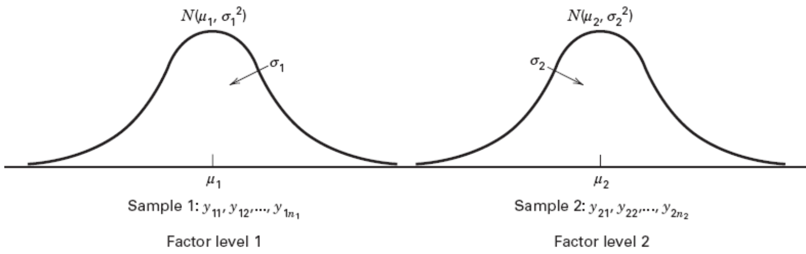
2.4.1. Hypothesis testing: Two-Sample *t*-Test

- Sample 1 (from population 1)

$$y_{11}, \dots, y_{1,n1}$$

- Sample 2 (from population 2)

$$y_{21}, \dots, y_{2,n2}$$



- Sample statistics (mean and variance)

$$\bar{y}_1, \bar{y}_2, s_1^2, s_2^2$$

- Assume: populations are Normal and Independently distributed with constant variance

$$y_{1j} \sim NID(\mu_1, \sigma^2)$$
$$y_{2j} \sim NID(\mu_2, \sigma^2)$$

- Test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

2.4.1. Hypothesis testing: Two-Sample *t*-Test

- Use the test statistic

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

- S_p^2 is the pooled sample variance (assumption $\sigma_1^2 = \sigma_2^2$)

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

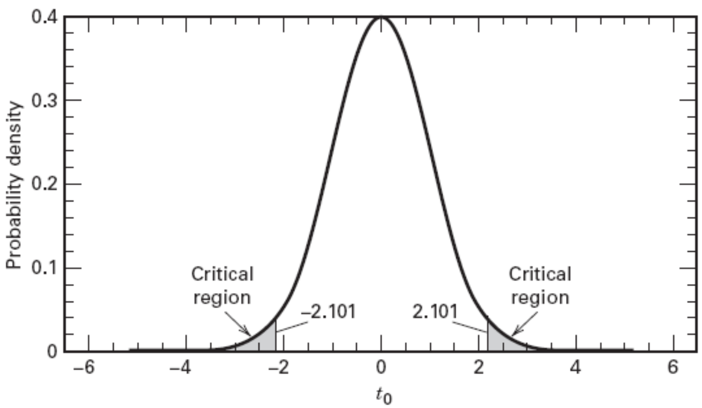
- Under H_0, the test statistic is distributed as

$$t_0 \sim t(n_1 + n_2 - 2)$$

  - How far apart the averages are, expressed in standard deviation units
  - Values of t_0 near 0 consistent with $H_0: \mu_1 = \mu_2$

One-sided alternative hypothesis

- H_0 rejected if $|t_0| > t_{\alpha/2,n_1+n_2-2}$ at α level of significance
- If we want to test H_1: $\mu_1 > \mu_2$
  - Reject H_0 only if $t_0 > t_{\alpha,n_1+n_2-2}$
- If we want to test H_1: $\mu_1 < \mu_2$
  - Reject H_0 only if $t_0 < -t_{\alpha,n_1+n_2-2}$



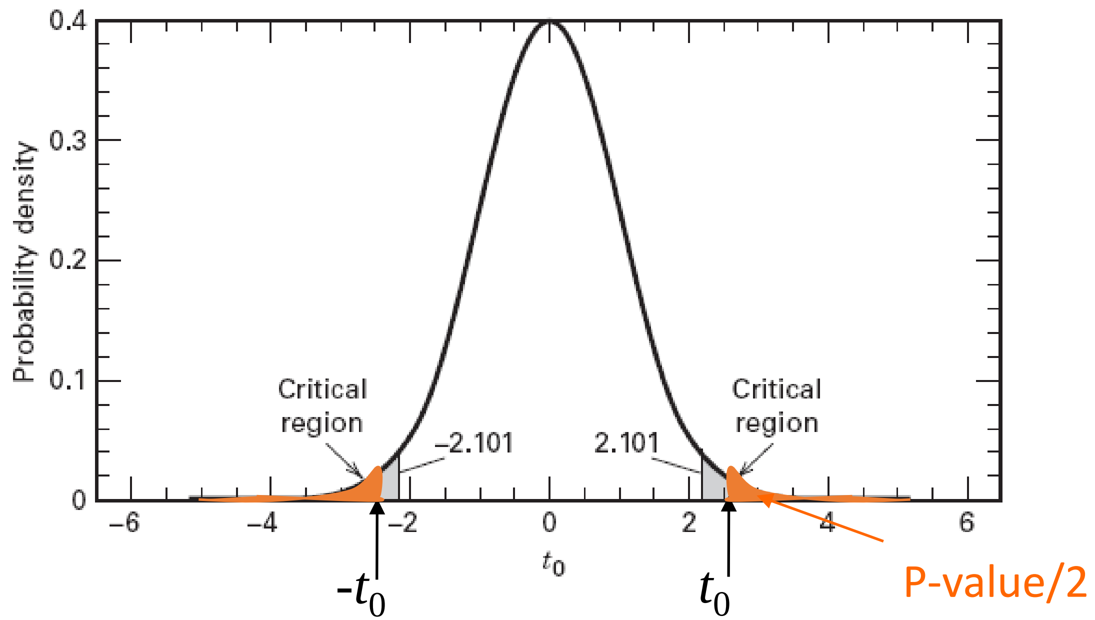
■ **FIGURE 2.10** The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$

For H_1: $\mu_1 \neq \mu_2$
Use $t_{0.025,18} = 2.101$
Reject H_0 if $|t_0| > 2.101$

For H_1: $\mu_1 > \mu_2$
Use $t_{0.05,18} = 1.734$
Reject H_0 if $t_0 > 1.734$

P-value approach to hypothesis testing

- Comparing t_0 to $t_{\alpha/2}$ is the fixed significance level testing
  - Does not tell whether t_0 is barely in the critical region or it was far into critical region
- *P*-value: area in the tails of the t-distribution beyond $-t_0$ and beyond t_0
  - Measure of how unusual the value of t_0 is, given H0 true
  - If $p < \alpha$ then Reject H_0 at α level of significance

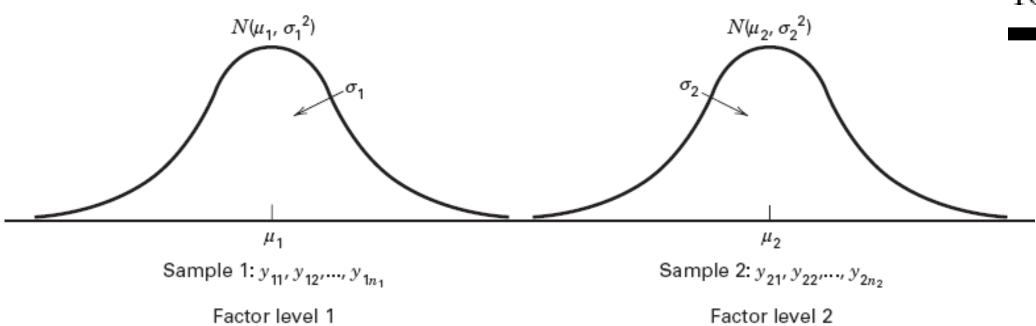


Example: Tension bond strength data (p. 26)

- Study whether adding a polymer to a cement impacts the strength.
- Test equality of modified mortar mean strength (μ_1) = unmodified mortar mean strength (μ_2)
- Test $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27



Summary Statistics

Formulation 1

"New recipe"

$$\bar{y}_1 = 16.76$$

$$S_1^2 = 0.100$$

$$S_1 = 0.316$$

$$n_1 = 10$$

Formulation 2

"Original recipe"

$$\bar{y}_2 = 17.04$$

$$S_2^2 = 0.061$$

$$S_2 = 0.248$$

$$n_2 = 10$$

Two-Sample *t*-Test

- Sample standard deviations are reasonably similar
  - Use pooled sample variance in the *t* statistic

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081$$

$$S_p = 0.284$$

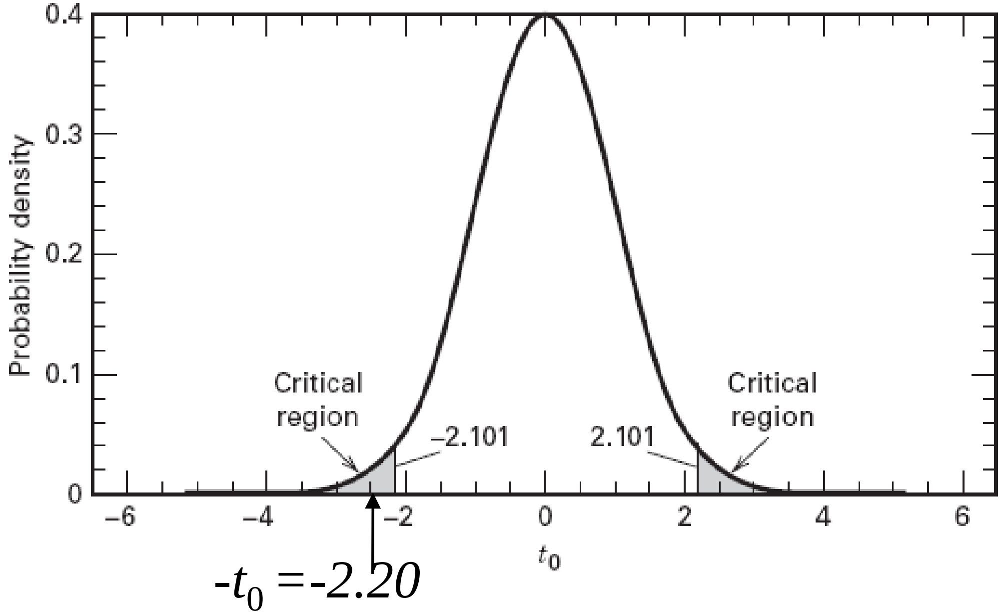
$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.20$$

The two sample means are a little over two standard deviations apart

Is this a "large" difference?

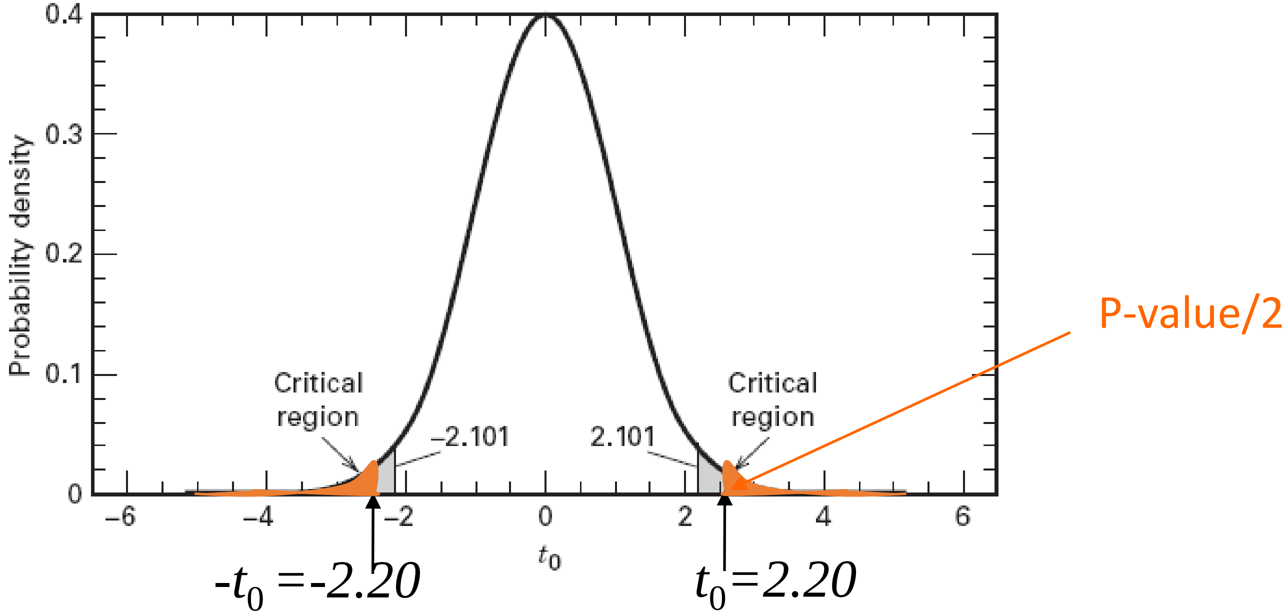
  - Use level of significance $\alpha = 0.05$
  - $t(\alpha/2, n_1 + n_2 - 2) = t(0.025{,}18) = 2.101$

Fixed Significance Level Approach



- A value of t_0 between -2.101 and 2.101 consistent with equality of means
- Since $t_0 = -2.20 < -2.101$, reject H_0
  - conclude mean tension bond strengths of two formulations are different at 0.05 level of significance

P-value approach



- *P*-value of observed test statistic $t_0 = -2.20$ is p = 0.042
- Need software to find exact *p*-value
- From Table II, we can find approximately $0.05 < p < 0.02$
- The hypothesis $H_0: \mu_1 = \mu_2$ rejected at any level of significance $\alpha \geq 0.042$

- Using t table to approximate p-value

II Percentage Points of the t Distribution[a]

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = Degrees of freedom.

[a]Adapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

Review last class: Simple comparative experiments

- Test $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$
- Sample 1 (from population 1)

$$y_{11}, \cdots, y_{1,n1}$$

- Sample 2 (from population 2)

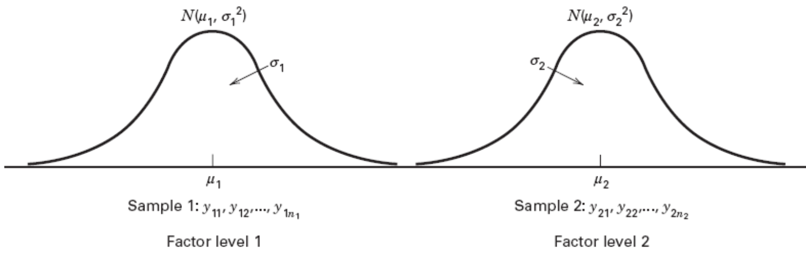
$$y_{21}, \cdots, y_{2,n2}$$

- Sample statistics (mean and variance)

$$\bar{y}_1, \bar{y}_2, s_1^2, s_2^2$$

- Assume: populations are Normal and Independently distributed with constant variance

$$y_{1j} \sim NID(\mu_1, \sigma^2)$$
$$y_{2j} \sim NID(\mu_2, \sigma^2)$$



Review last class: The Two-Sample t-Test on cement bond strength experiment

- Study whether adding a polymer to a cement impacts the bonding strength.
- Make 10 samples with original formulation and 10 samples with modified formulation. Measure strength.
- Test equality of modified mortar mean strength (μ_1) = unmodified mortar mean strength (μ_2)
- Two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Review last class: The Two-Sample *t*-Test

- Use the sample means to draw inference about the difference between two population means

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081$$

$$S_p = 0.284$$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.20$$

The two sample means are a little over two standard deviations apart

Is this a "large" difference?

- $t_{\alpha/2, n_1+n_2-2} = t_{0.025,18} = 2.101$. Since $|t_0| > 2.101$, the observed difference is significant→ the strengths of two formulations are different

2.4.2. Confidence intervals for two-sample test

- Hypothesis testing only tells whether μ_1 and μ_2 differ
  - It does not tell how much they differ
  - confidence interval on difference between means $\mu_1 - \mu_2$
- An interval $L \leq \mu_1 - \mu_2 \leq U$ within which the true difference $\mu_1 - \mu_2$ lies with high confidence

2.4.2. Confidence intervals for two-sample test

- Hypothesis testing only tells whether μ_1 and μ_2 differ
  - It does not tell how much they differ
  - confidence interval on difference between means $\mu_1 - \mu_2$
- An interval $L \leq \mu_1 - \mu_2 \leq U$ within which the true difference $\mu_1 - \mu_2$ lies with high confidence
  - A $100(1-\alpha)\%$ confidence interval on $\mu_1 - \mu_2$

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\leq \mu_1 - \mu_2 \leq$$

$$\bar{y}_1 - \bar{y}_2 + t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Cement bond strength experiment

- Estimated difference between mean bond strengths :
$$\bar{y}_1 - \bar{y}_2 = 16.76 - 17.04 = -0.278$$
- 95% confidence interval for the true difference between the means

Cement bond strength experiment

- Estimated difference between mean bond strengths :

$$\bar{y}_1 - \bar{y}_2 = 16.76 - 17.04 = -0.278$$

- 95% confidence interval for the true difference between the means

$$16.76 - 17.04 - (2.101)0.284\sqrt{\frac{1}{10} + \frac{1}{10}} \leq \mu_1 - \mu_2 \leq$$

$$16.76 - 17.04 + (2.101)0.284\sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$-0.545 \leq \mu_1 - \mu_2 \leq -0.111$$

- Since 0 is not included in the 95% confidence interval, we reject H_0 at 5% significance level

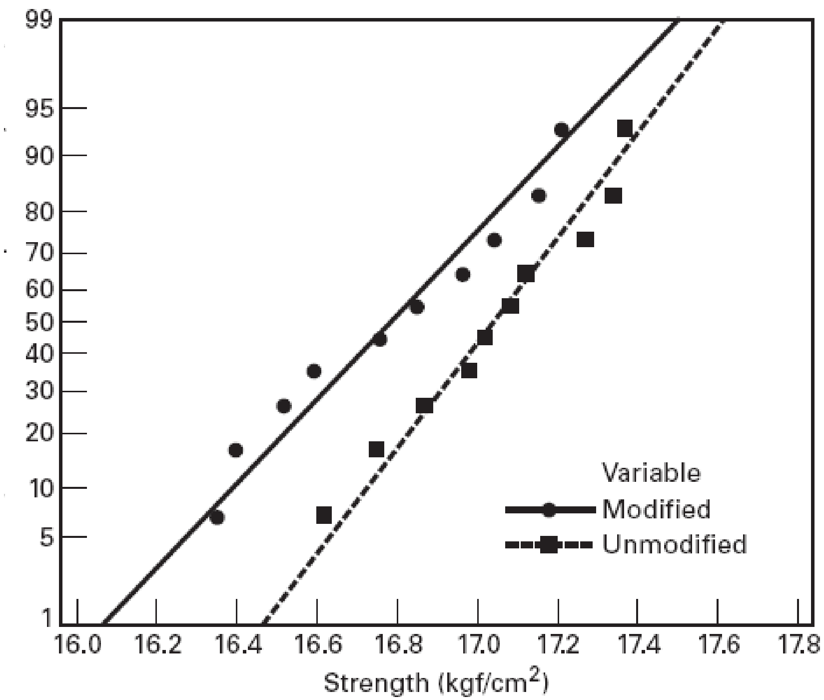
Checking Assumptions – Normal Probability Plot

- Assumption in t-test: Samples are from independent Normal populations with equal standard deviation

$$y_{11}, \ldots, y_{1,n1} \sim NID(\mu_1, \sigma^2)$$
$$y_{21}, \ldots, y_{2,n2} \sim NID(\mu_2, \sigma^2)$$

- If assumptions violated, performance of t-test will be affected
  - significance level and ability to detect differences between means will be different than advertised
  - Variance stabilizing transformations (Chapter 3) or nonparametric tests for non-normality

- To check normality, "Normal Probability Plot" of the data

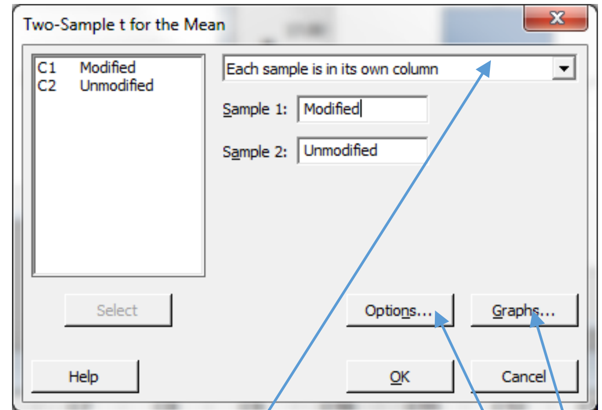


Minitab: Two-Sample *t*-Test on cement bond strength experiment

- Enter the data in two columns
  - Strength of samples under two conditions

Minitab: Two-Sample *t*-Test on cement bond strength experiment

Minitab: Two-Sample *t*-Test on cement bond strength experiment

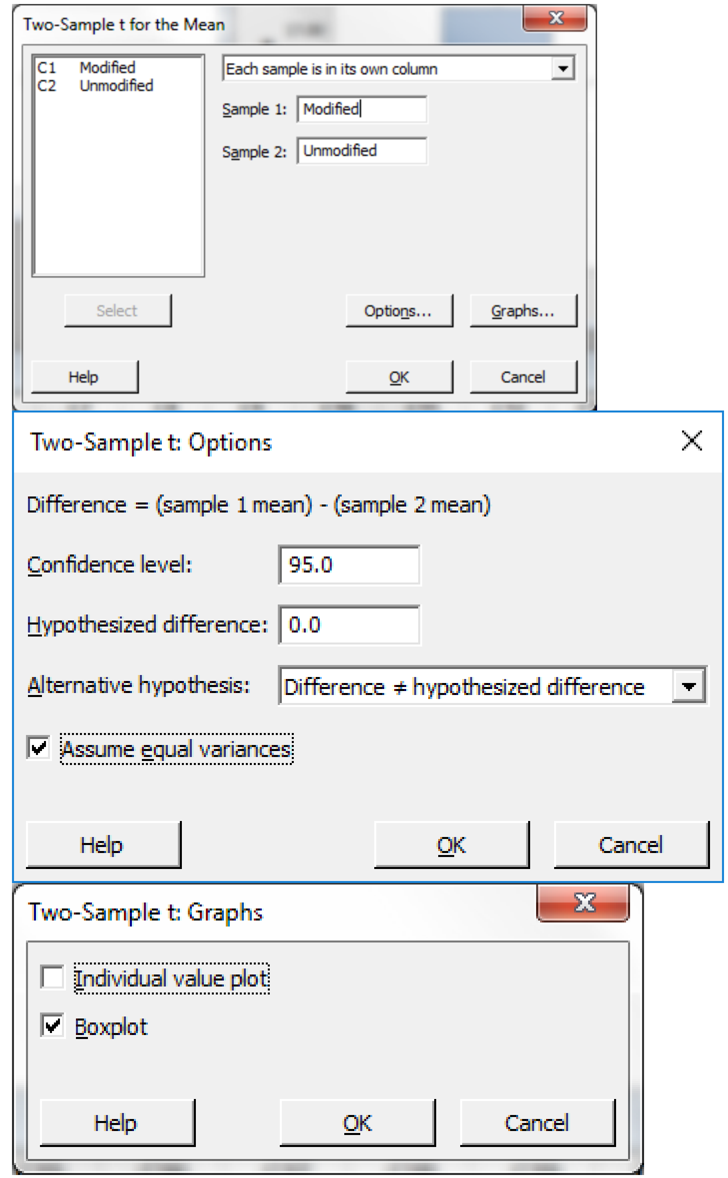


Sample entry options

- Each sample in its own column
- Samples are in one column
- Summarized data

Click OK to conduct t-test with defaults
Or
Click Options and Graphs to make changes

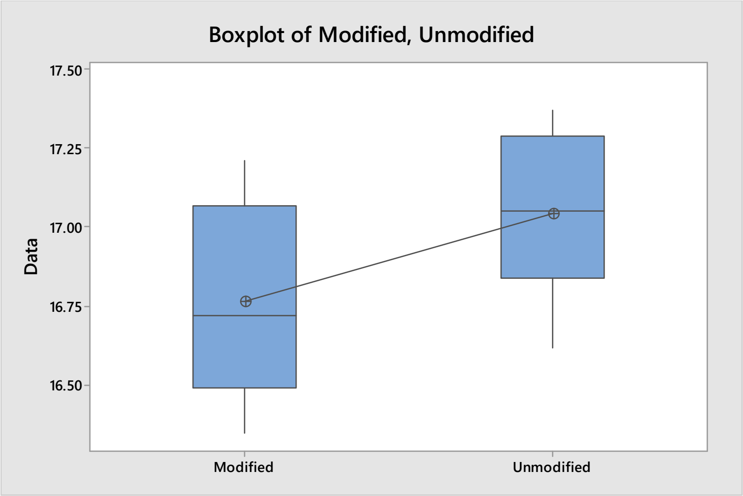
Minitab: Two-Sample *t*-Test on cement bond strength experiment



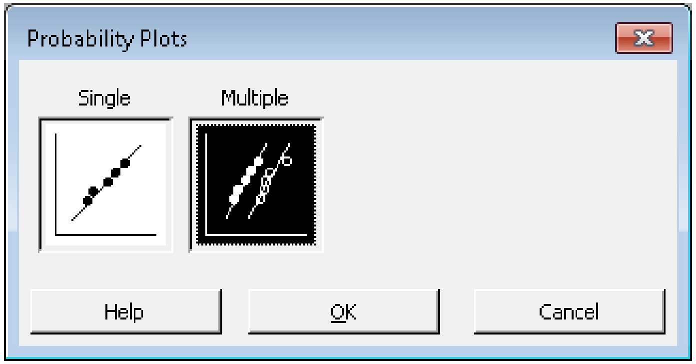
Minitab: Two-Sample *t*-Test on cement bond strength experiment

Two-Sample T-Test and CI: Modified, Unmodified

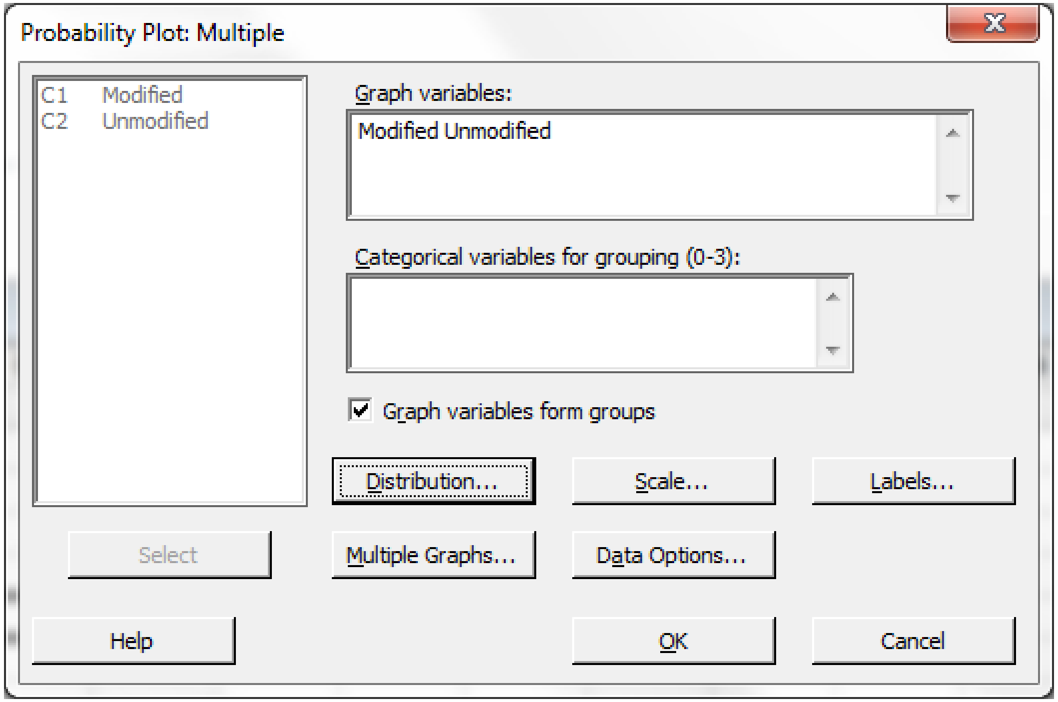
```
Two-sample T for Modified vs Unmodified

              N    Mean  StDev  SE Mean
Modified     10  16.764  0.316     0.10
Unmodified   10  17.042  0.248    0.078


Difference = μ (Modified) - μ (Unmodified)
Estimate for difference:  -0.278
95% CI for difference:  (-0.545, -0.011)
T-Test of difference = 0 (vs ≠): T-Value = -2.19  P-Value = 0.042  DF = 18
Both use Pooled StDev = 0.2843
```



Probability plotting: cement bond strength experiment

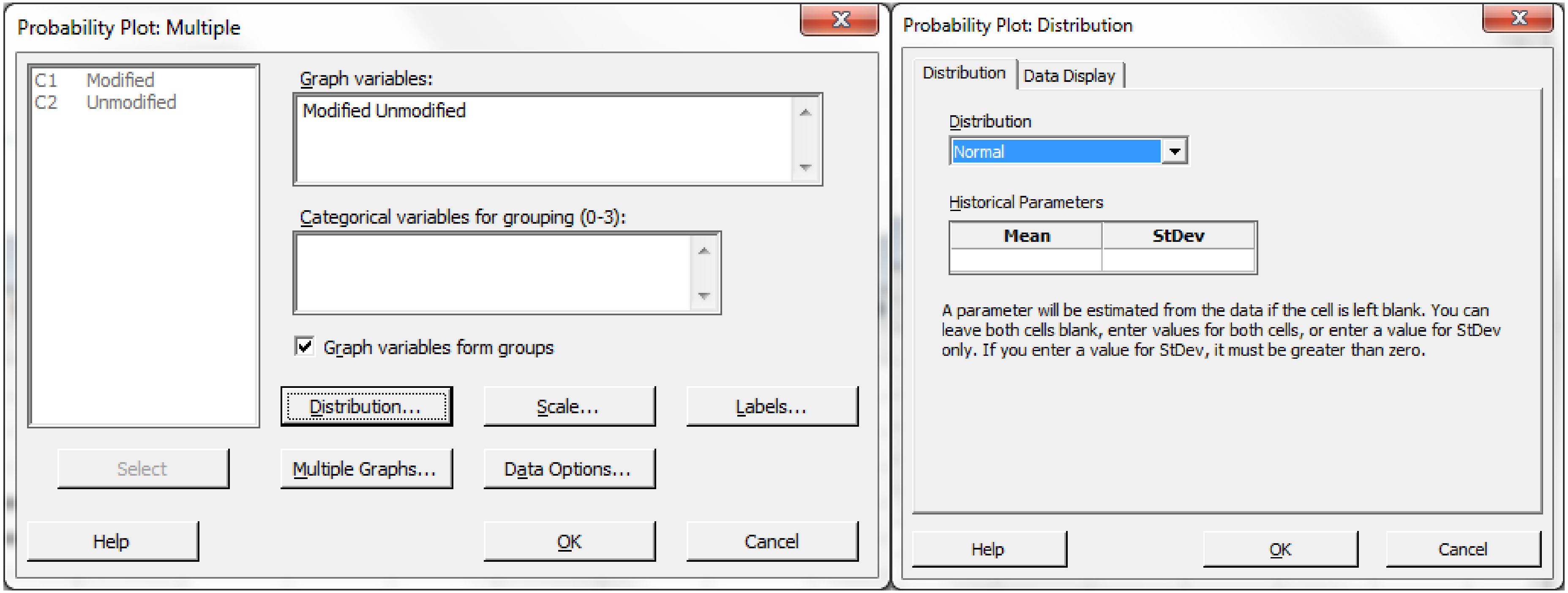


Probability plotting: cement bond strength experiment



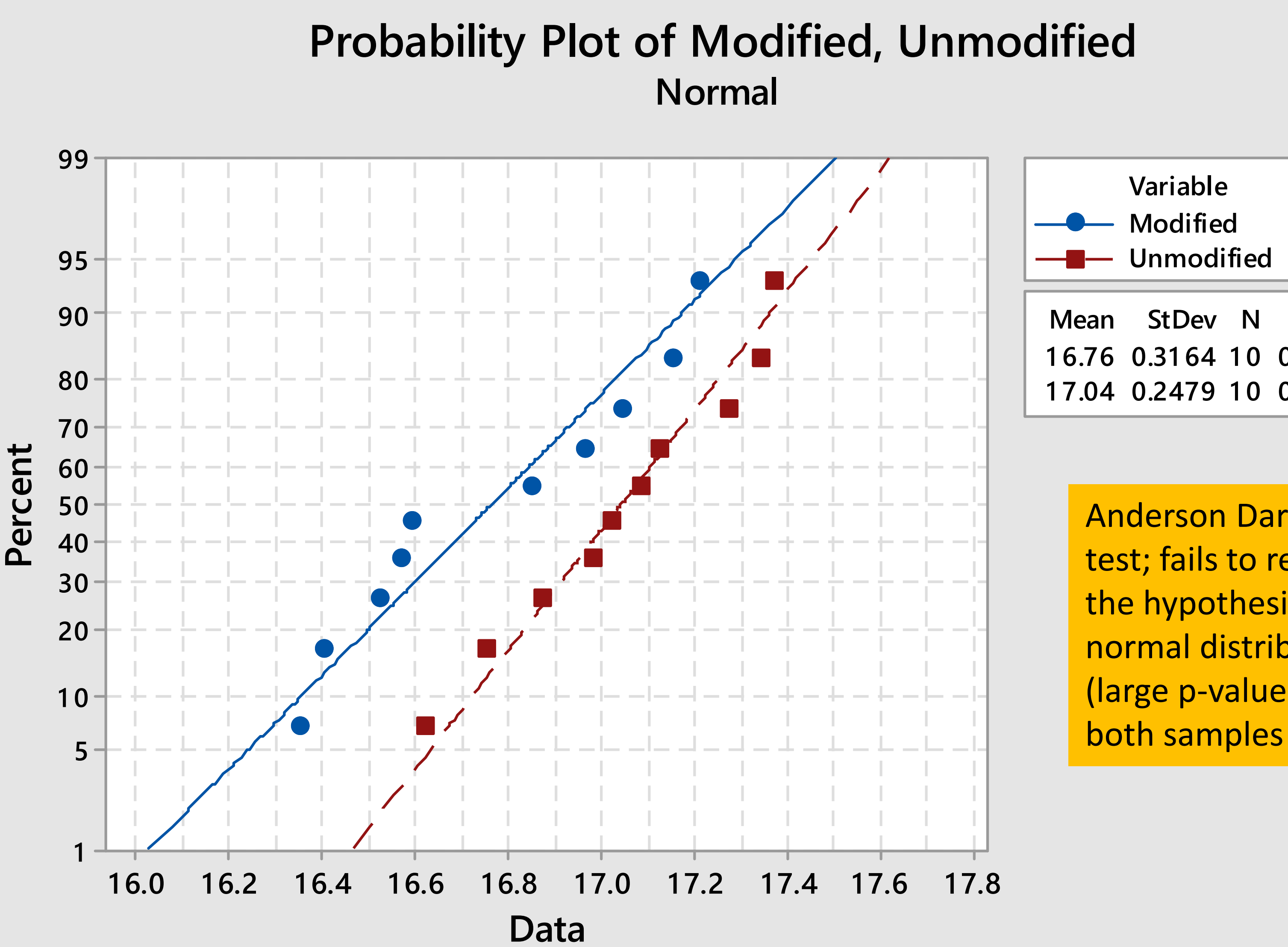
- Choose the Columns for the samples to be analyzed
- Click OK for probability plot with Normal
- Click Distribution to change the distribution

Probability plotting: cement bond strength experiment



- Choose the Columns for the samples to be analyzed
- Click OK for probability plot with Normal
- Click Distribution to change the distribution

Probability plotting: cement bond strength experiment



Mean	StDev	N	AD	P
16.76	0.3164	10	0.353	0.388
17.04	0.2479	10	0.170	0.905

Anderson Darling test; fails to reject the hypothesized normal distribution (large p-value) for both samples

2.4.3. Choice of sample size (n_1 and n_2)

- Hypothesis test for inference between two means

$$H_0: \mu_1 = \mu_2 \text{ vs } H_1: \mu_1 \neq \mu_2$$

  - Reject H_0 if $|t_0| > t_{\alpha, n_1+n_2-2}$
- Two kinds of errors may be committed

$$\alpha = Pr(\text{type I error}) = Pr(H_0 \text{ is rejected when it is true })$$

$$\beta = Pr(\text{type II error }) = Pr(H_0 \text{ is not rejected when false})$$

- Specify α → determines the critical region
  - α: level of significance of the test

2.4.3. Choice of sample size (n_1 and n_2)

- Hypothesis test for inference between two means

$$H_0: \mu_1 = \mu_2 \text{ vs } H_1: \mu_1 \neq \mu_2$$

  - Reject H_0 if $|t_0| > t_{\alpha, n_1+n_2-2}$
- Two kinds of errors may be committed

$$\alpha = Pr(\text{type I error}) = Pr(H_0 \text{ is rejected when it is true })$$
$$\beta = Pr(\text{type II error }) = Pr(H_0 \text{ is not rejected when false})$$

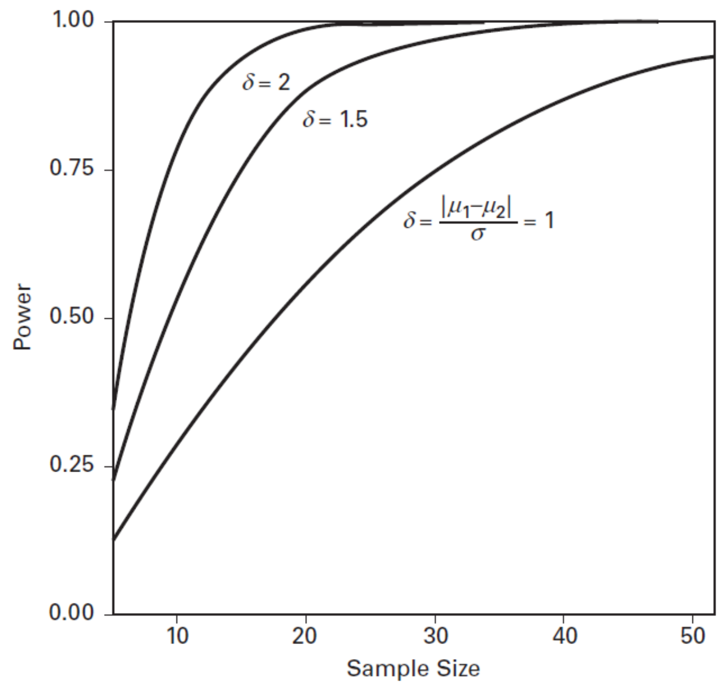
- Specify α → determines the critical region
  - α: level of significance of the test
- Choose sample size ($n = n_1 = n_2$, balanced design) so that β is suitably small
  - Or Power $= 1 - \beta$ is suitably large
  - Choose sample size to have large power (probability) of discriminating μ_1 from μ_2

2.4.3. Choice of sample size (n_1 and n_2)

- Power depends on sample size n, difference between means $|\mu_1 - \mu_2|$ and experimental error σ (For α fixed)
- Procedure: For a specified $|\mu_1 - \mu_2|$ that is of practical interest and assumed worst case σ, choose n so that power is large
  - if the difference in means is at least this large, we can detect it with a high probability
  - Plot power against n or scaled difference
    $$\delta = |\mu_1 - \mu_2|/\sigma$$
- Power curve
  - A graph of Power versus δ (given n)
  - A graph of Power versus n (given δ)

Power Curves for the Two-Sample t-Test assuming equal variances and $\alpha = 0.05$.

- Power versus $2n$ (given δ)
  - From JMP



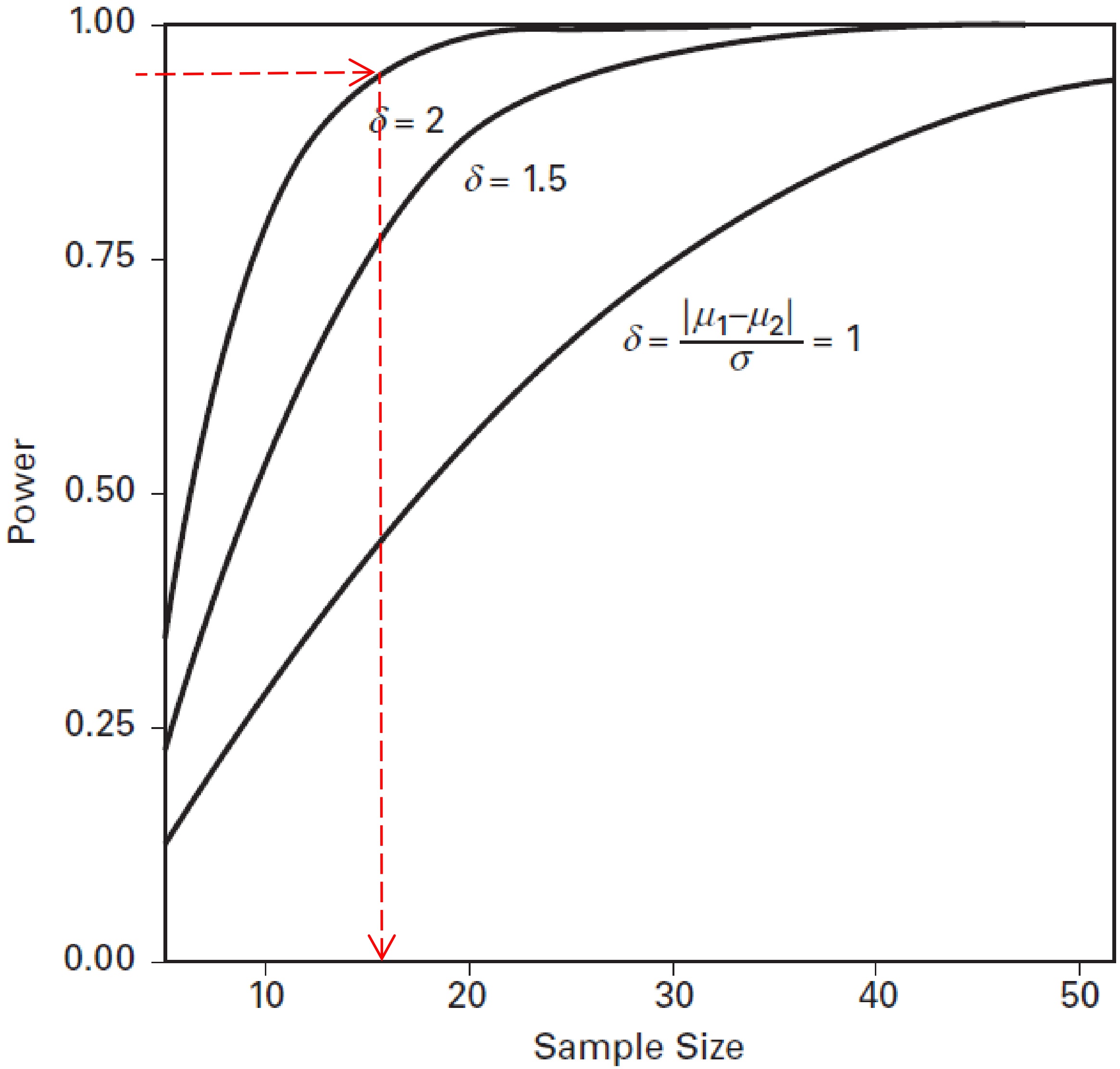
- Power versus $\mu_1 - \mu_2$ (given n)
  - From Minitab

Cement bond strength experiment: Choose sample size for Two-Sample *t*-Test

- Suppose that a difference in mean strength of $|\mu_1 - \mu_2| = 0.5$ kgf/cm2 has practical impact on the use of the mortar
- Also assume that experimental error σ is unlikely to be larger than 0.25 kgf/cm2.
- Choose sample size so that Power is at least 0.95
  - detect difference $\delta = \frac{0.5}{0.25} = 2$ with power =1- β = 0.95

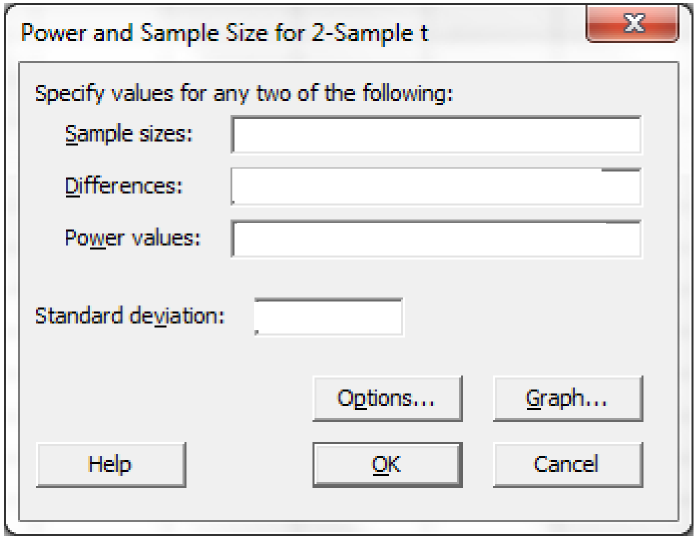
For power = 0.95 and $\delta = 2$ required total sample size on horizontal axis is 16.

Thus $n = \frac{16}{2} = 8$



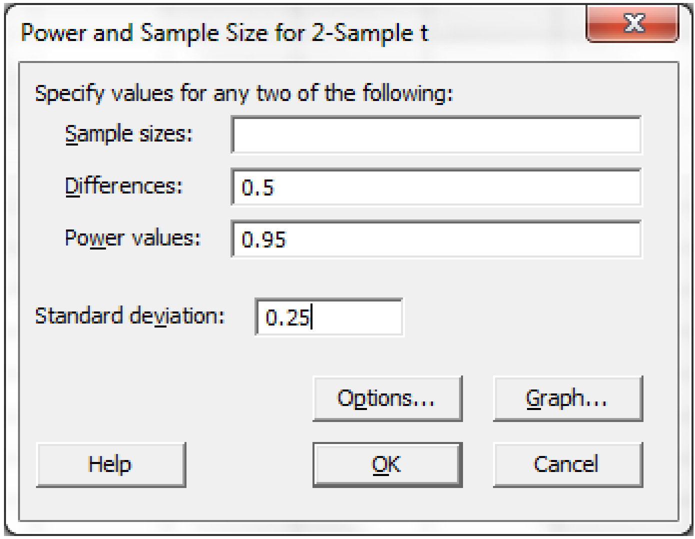
Sample size and power in Minitab

- For sample size calculation spreadsheet data is not used
- Find sample size for a specified difference in means = 0.5 and power = 0.95
  - Assumed standard deviation = 0.25



Sample size and power in Minitab

- For sample size calculation spreadsheet data is not used
- Find sample size for a specified difference in means = 0.5 and power = 0.95
  - Assumed standard deviation = 0.25

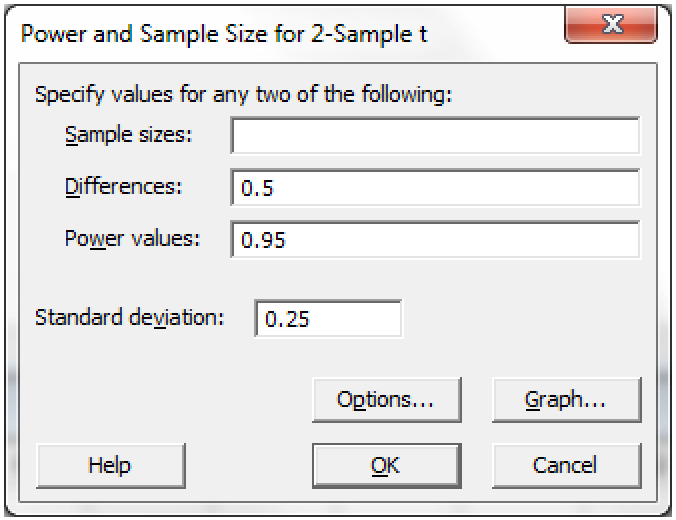


Sample size and power in Minitab

- Minitab suggests to use at least $n_1 = n_2 = 8$ samples

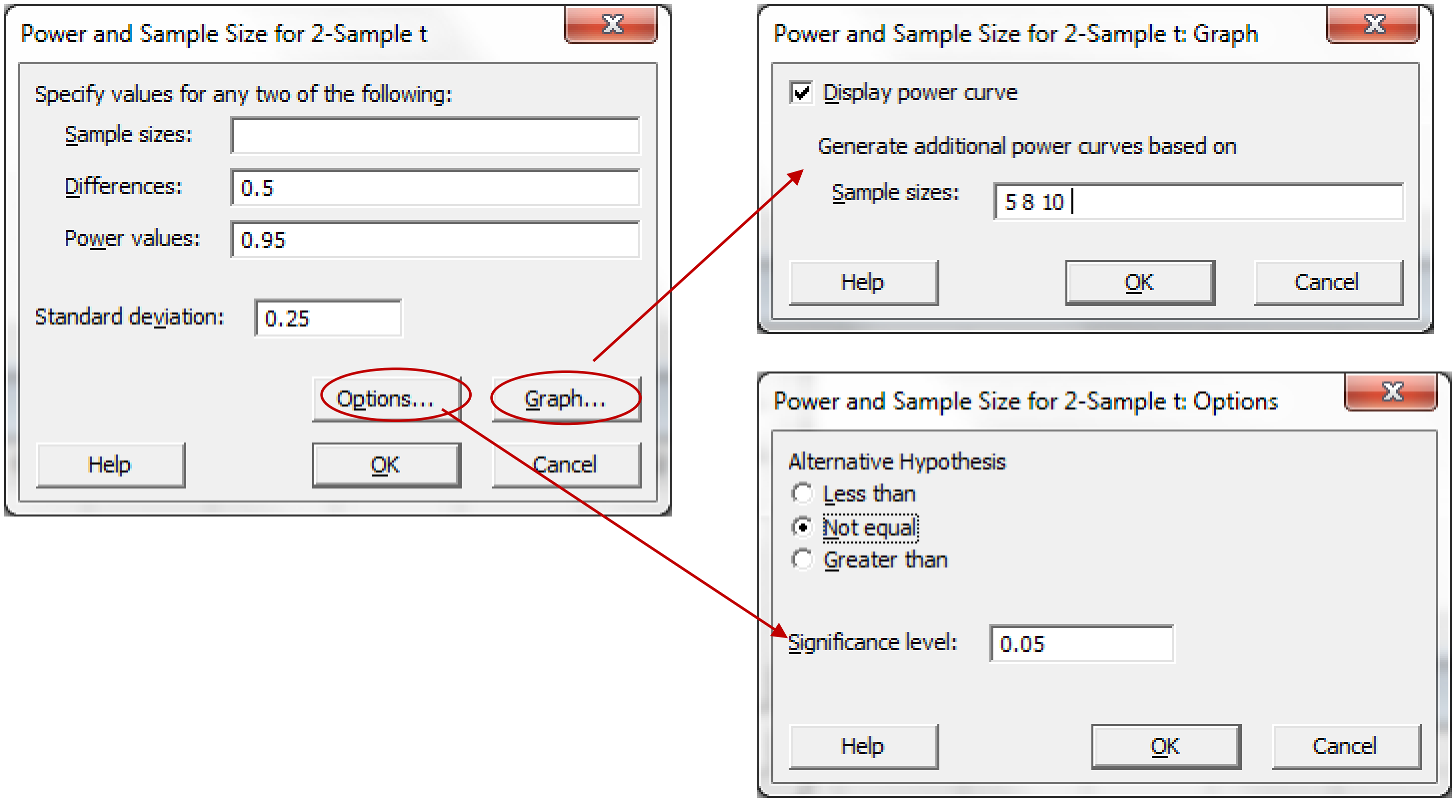
```
Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 0.25


            Sample  Target
Difference    Size   Power  Actual Power
       0.5       8    0.95      0.960221

The sample size is for each group.
```

Generate power curves for different sample sizes

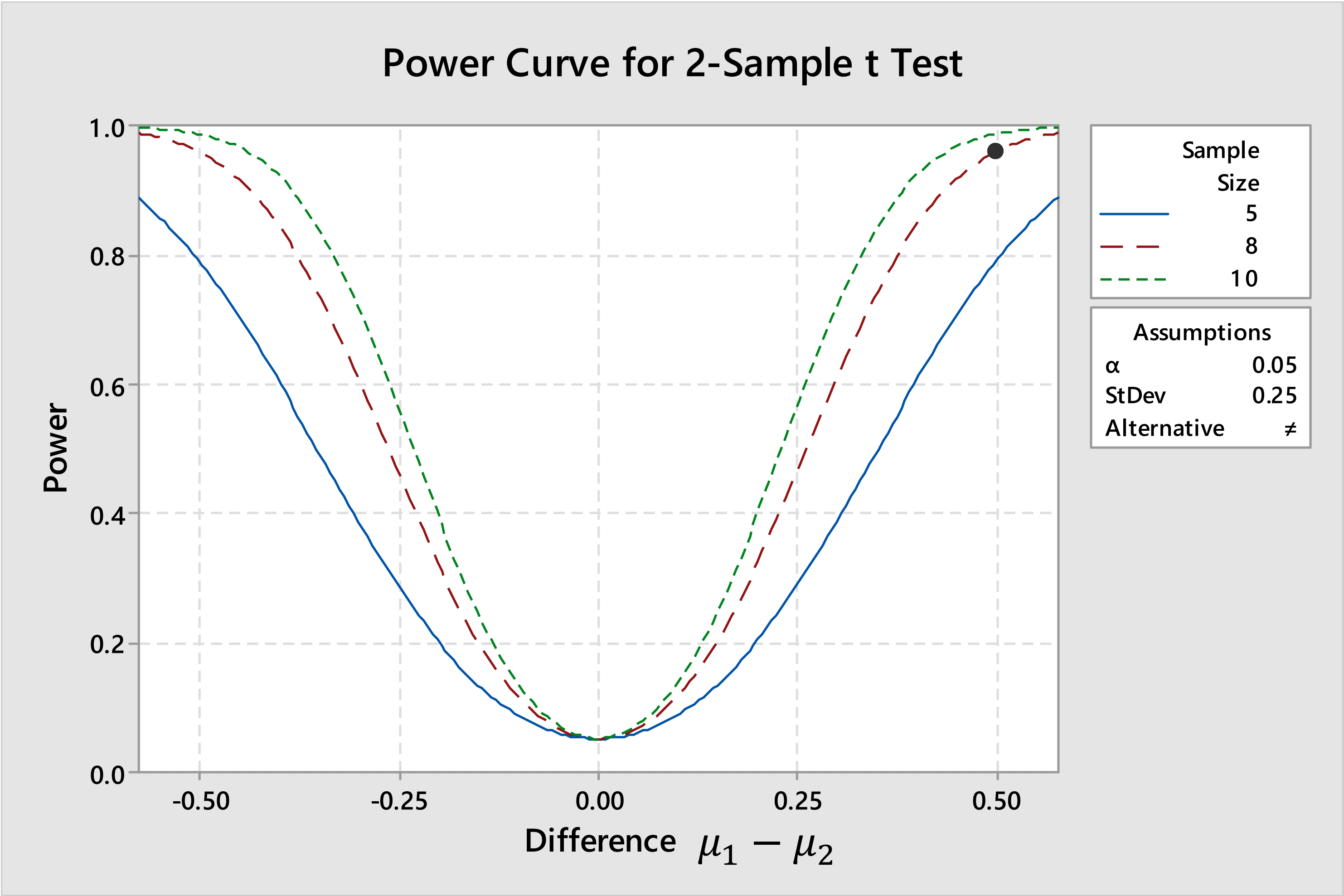


- Click OK to find sample size to satisfy power=0.95 with default options
- Click Options and Graph to make changes or plot the Power Curve

Generate power curves for different sample sizes



- A graph of Power versus $\mu_1 - \mu_2$ (given n)



2.5. Paired Comparison Problem

- In simple comparative experiment we can improve the precision of treatment comparisons by making the comparisons between “matched pairs” of experimental unit
  - Test difference between means of two independent populations (two-sample t-test, or Completely Randomized Design, CRD)
  - Test difference between means of two paired populations (paired t-test, or Randomized Complete Block Design, RCBD)

Example: hardness testing

- Compare hardness measurements of two different tips on 10 sample coupons
  - Coupon is an experimental unit
  - Tip is the treatment

■ **TABLE 2.5**
Data for the Hardness Testing Experiment

Specimen	Tip 1	Tip 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

Two experimental strategies

- CRD will apply 2 tips randomly on 20 different coupons
  - the variations in coupon properties may affect the comparison
- RCBD: To block this variation, apply tip1 and tip2 on the same coupon (assuming each coupon large enough to conduct 2 tests)
  - Compare the hardness values on the same coupon
  - need 10 large coupons
  - Find difference in hardness values on each coupon

Two experimental strategies

- CRD will apply 2 tips randomly on 20 different coupons
  - ❑ the variations in coupon properties may affect the comparison
- RCBD: To block this variation, apply tip1 and tip2 on the same coupon (assuming each coupon large enough to conduct 2 tests)
  - ❑ Compare the hardness values on the same coupon
  - ❑ need 10 large coupons
  - ❑ Find difference in hardness values on each coupon
- Test on relatively homogeneous experimental unit (Experimental error in RCBD is smaller than CRD)
  - ❑ Test will have higher power to detect a given difference between means

Assume the experiment is a RCBD: To compare the treatments use paired t-test

- Measurement from Population 1: $y_1 \sim N(\mu_1, \sigma^2)$
- Measurement from Population 2: $y_2 \sim N(\mu_2, \sigma^2)$
- Difference between two measurements:

$$d = y_1 - y_2$$

$$d \sim N\left(\mu_d, \sigma_d^2\right)$$

- Testing $H_0\colon \mu_1 - \mu_2 = 0$ is equivalent to testing

$$H_0\colon \mu_d = 0$$

- One-sample t-test on the paired differences

Assume the experiment is a RCBD: To compare the treatments use paired t-test

- Observed values
  - with treatment 1: $y_{11}, y_{12}, \dots, y_{1,n}$
  - with treatment 2: $y_{21}, y_{22}, \dots, y_{2,n}$
- Observed matched differences

$$
\begin{gathered}
d_1 = y_{11} - y_{21}, \\
d_2 = y_{12} - y_{22}, \\
\dots, \\
d_n = y_{1,n} - y_{2,n}
\end{gathered}
$$

- Sample mean and variance of the differences

$$\bar{d}, \text{ and } s_d$$

- Use test statistic

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

- Under H_0, t_0 is distributed as t_{n-1}. Reject H_0 if $|t_0| > t_{\frac{\alpha}{2},n-1}$

Example: hardness testing

■ **TABLE 2.5**

Data for the Hardness Testing Experiment

Specimen	Tip 1	Tip 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

- Paired differences

$$d_1 = 7 - 6 = 1 \qquad d_6 = 3 - 2 = 1$$
$$d_2 = 3 - 3 = 0 \qquad d_7 = 2 - 4 = -2$$
$$d_3 = 3 - 5 = -2 \qquad d_8 = 9 - 9 = 0$$
$$d_4 = 4 - 3 = 1 \qquad d_9 = 5 - 4 = 1$$
$$d_5 = 8 - 8 = 0 \qquad d_{10} = 4 - 5 = -1$$

- Mean and standard deviation of paired differences

$$\bar{d} = \frac{1}{n}\sum_{j=1}^{n} d_j = \frac{1}{10}(-1) = -0.10$$

$$S_d = \left[\frac{\sum_{j=1}^{n} d_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} d_j\right)^2}{n-1}\right]^{1/2} = \left[\frac{13 - \frac{1}{10}(-1)^2}{10-1}\right]^{1/2} = 1.20$$

- Paired -t test statistic

$$t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{-0.10}{1.20/\sqrt{10}} = -0.26$$

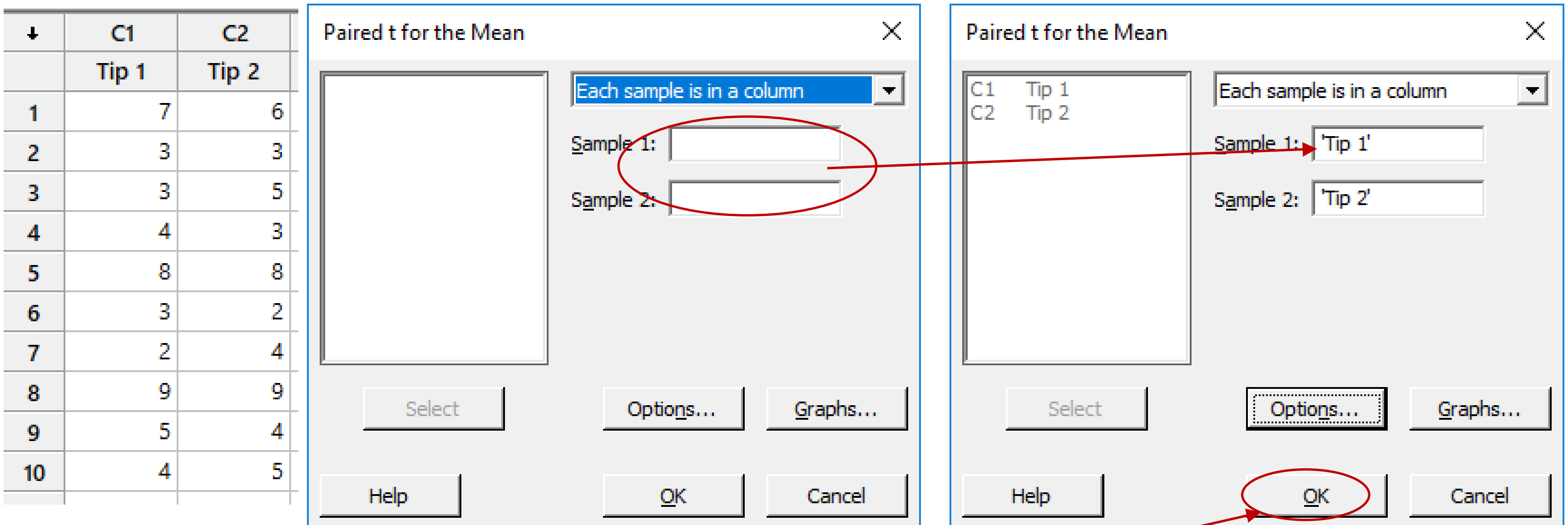
- Decision: $t_{0.025,9} = 2.262$. Since $|t_0| < 2.262$ we cannot reject H_0

Paired t- test in Minitab: Hardness test

- Create the spreadsheet
- Choose: Basic Statistics>Paired t

↓	C1	C2
	Tip 1	Tip 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

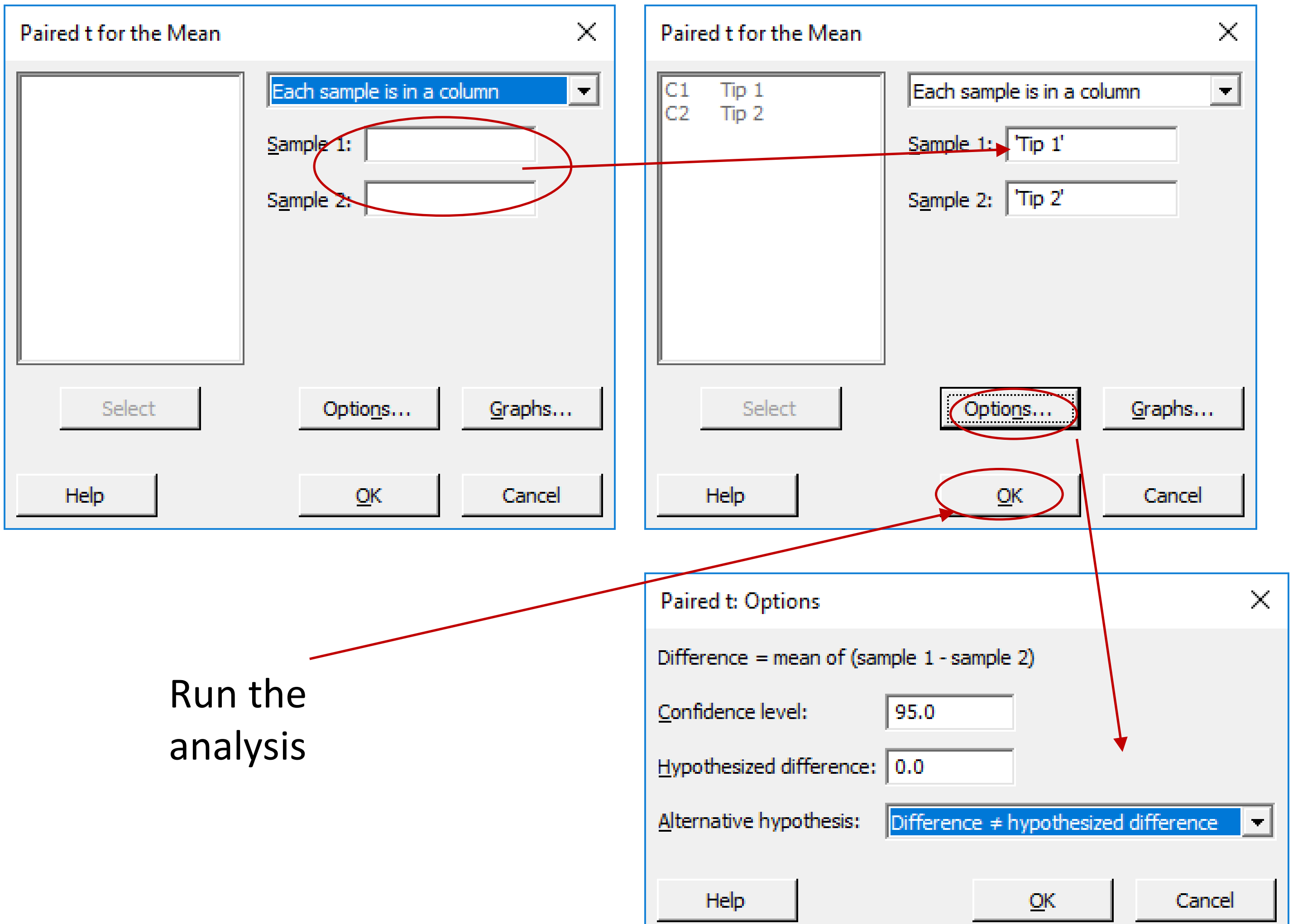
- each sample data is in a column



Run the
analysis

- each sample data is in a column

↓	C1	C2
	Tip 1	Tip 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5



Run the analysis

2.5. Paired Comparison Problem

- Compare $t_0 = -0.26$ to $t(n-1, \alpha/2) = t(9{,}0.025) = 2.262$

```
Paired T-Test and CI: Tip 1, Tip 2

Paired T for Tip 1 - Tip 2

              N       Mean     StDev   SE Mean
Tip 1        10    4.80000   2.39444   0.75719
Tip 2        10    4.90000   2.23358   0.70632
Difference   10  -0.100000  1.197219  0.378594


95% CI for mean difference: (-0.956439, 0.756439)
T-Test of mean difference = 0 (vs not = 0): T-Value =
   -0.26  P-Value = 0.798
```

2.5. Paired Comparison Problem

- Independent (two-sample t) analysis

```
Two-Sample T-Test and CI: Tip 1, Tip 2

Two-sample T for Tip 1 vs Tip 2

        N   Mean   StDev   SE Mean
Tip 1  10   4.80    2.39      0.76
Tip 2  10   4.90    2.23      0.71

Difference = mu (Tip 1) - mu (Tip 2)
Estimate for difference:  -0.100000
95% CI for difference:  (-2.275466, 2.075466)
T-Test of difference = 0 (vs not =): T-Value =
  -0.10  P-Value = 0.924  DF = 18
Both use Pooled StDev = 2.3154
```

Paired-t test has higher precision for comparing two treatment means than two-sample t

- **much shorter CI for mean difference**
- **smaller P-value (more power)**

2.6 Inferences on variances

- In two sample tests we assume equal variance of two populations
- Hypothesis tests on variances

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

- Use test statistic

$$F_0 = \frac{s_1^2}{s_2^2}$$

- Which follows $F(n_1 - 1, n_2 - 1)$ when H_0 is true
  - Decision rule: Reject H_0 at α level if

$$F_0 > F(\alpha/2, n_1 - 1, n_2 - 1) \text{ or}$$
$$F_0 < F(1 - \alpha/2, n_1 - 1, n_2 - 1)$$

Example: Cement bond strength data

↓	C1	C2
	Modified	Unmodified
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

- Test at $\alpha = 0.05$

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_0: \sigma_1^2 \neq \sigma_2^2$$

- Sample variances
  - ❑ $S_1^2 = 0.100$
  - ❑ $S_2^2 = 0.061$
- F test statistic, and critical value
  - ❑ $F_0 = S_1^2 / S_2^2 = 1.667$
  - ❑ $F(0.05,9,9) = 3.18$
  - ❑ $F(0.95,9,9) = \frac{1}{F(0.05,9,9)} = \frac{1}{3.18} = 0.314$
- No evidence that variances are different
  - ❑ Fail to reject H_0

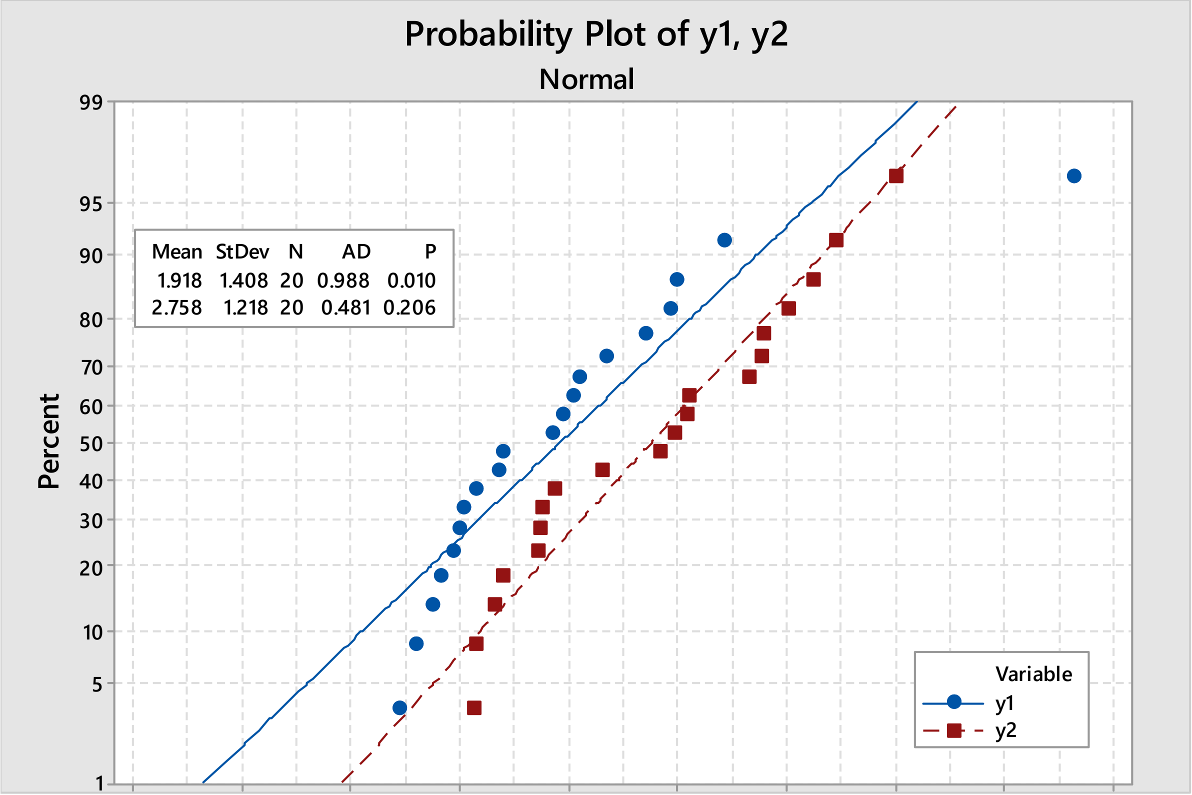
Non-parametric two sample t-tests

- For two-sample comparisons, if normal distribution assumption is not satisfied non-parametric tests can be used
  - Non-parametric test → distribution free tests
- two sample test
  - use Mann-Whitney-Wilcoxon rank sum test
- paired sample tests
  - use Wilcoxon signed rank test.

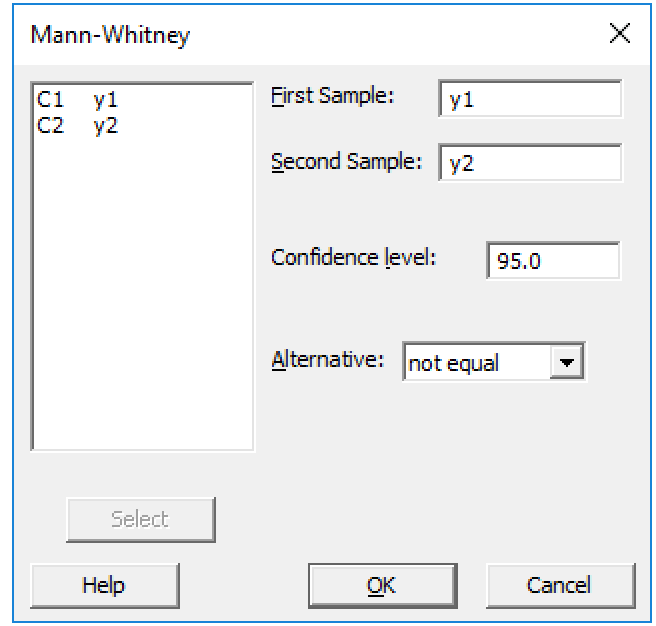
Example – data from two nonnormal populations

- Test whether the means of the two populations are equal
$$H_0 : \mu_1 = \mu_2$$
- 20 data points from two populations. Normality assumption not satisfied in the two samples: use Mann-Whitney non-parametric test

↓	C1	C2
	y1	y2
1	1.36	3.64
2	0.82	1.31
3	2.98	1.74
4	2.34	2.96
5	3.42	1.39
6	0.60	4.02
7	1.03	1.12
8	2.93	1.73
9	6.62	3.76
10	0.75	1.71
11	1.14	1.87
12	1.00	2.82
13	2.09	1.15
14	2.04	2.30
15	1.94	3.11
16	0.93	4.25
17	2.70	5.00
18	1.38	3.07
19	1.84	4.44
20	0.44	3.78



Mann-Whitney test



Mann-Whitney Test and CI: y1, y2

```
     N  Median
y1  20   1.611
y2  20   2.891


Point estimate for η1 - η2 is -0.909
95.0 Percent CI for η1 - η2 is (-1.801,-0.220)
W = 320.0
Test of η1 = η2 vs η1 ≠ η2 is significant at 0.0155
```

Reject $H0\colon \mu_1 = \mu_2$ because *p*-value = 0.0155

Back-up

Degrees of freedom of a SS

The sum of squares

$$SS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

has $n - 1$ degrees of freedom because the SS has $n - 1$ independent elements.

Also

$$E\left[\frac{SS}{n-1}\right] = E[s^2] = \sigma^2$$

2.4.3. Confidence Intervals

- Hypothesis testing gives an objective answer to the question "Are the two means equal?"
- But it doesn't tell "how different" they are
  - In many systems we would already know $\mu_1 \neq \mu_2$
  - We are interested in a confidence interval on μ_1-μ_2
- **General form** of a confidence interval

$$L \leq \theta \leq U \text{ where } P(L \leq \theta \leq U) = 1 - \alpha$$

- The 100(1- α)% **confidence interval** on the difference in two means:

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2, n_1+n_2-2} S_p \sqrt{(1/n_1) + (1/n_2)} \leq \mu_1 - \mu_2 \leq \bar{y}_1 - \bar{y}_2 + t_{\alpha/2, n_1+n_2-2} S_p \sqrt{(1/n_1) + (1/n_2)}$$

2.4.3. Confidence Intervals

- The interval can be derived as follows.
- The statistic t is distributed as $t_{n1+n2-2}$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Therefore

$$P\left(-t_{\frac{\alpha}{2},n_1+n_2-2} \le \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{\frac{\alpha}{2},n_1+n_2-2}\right) = 1 - \alpha$$

- And the CI on μ_1-μ_2 can be obtained

Portland cement data

Formulation 1

"New recipe"

$$\bar{y}_1 = 16.76$$

$$S_1^2 = 0.100$$

$$S_1 = 0.316$$

$$n_1 = 10$$

Formulation 2

"Original recipe"

$$\bar{y}_2 = 17.04$$

$$S_2^2 = 0.061$$

$$S_2 = 0.248$$

$$n_2 = 10$$

$$\mathrm{S_p} = 0.284$$

$$t_{\alpha/2,n1+n2\text{-}2} = 2.101$$

The actual 95 percent confidence interval estimate for the difference in mean tension bond strength for the formulations of Portland cement mortar is found by substituting in Equation 2.30 as follows:

$$\begin{aligned}
16.76 - 17.04 - (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}} &\leq \mu_1 - \mu_2 \\
&\leq 16.76 - 17.04 + (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}} \\
-0.28 - 0.27 \leq \mu_1 - \mu_2 &\leq -0.28 + 0.27 \\
-0.55 \leq \mu_1 - \mu_2 &\leq -0.01
\end{aligned}$$

Thus, the 95 percent confidence interval estimate on the difference in means extends from -0.55 to -0.01 kgf/cm^2. Put another way, the confidence interval is $\mu_1 - \mu_2 = -0.28 \pm 0.27$ kgf/cm^2, or the difference in mean strengths is -0.28 kgf/cm^2, and the accuracy of this estimate is ± 0.27 kgf/cm^2. Note that because $\mu_1 - \mu_2 = 0$ is *not* included in this interval, the data do not support the hypothesis that $\mu_1 = \mu_2$ at the 5 percent level of significance (recall that the *P*-value for the two-sample *t*-test was 0.042, just slightly less than 0.05). It is likely that the mean strength of the unmodified formulation exceeds the mean strength of the modified formulation. Notice from Table 2.2 that both Minitab and JMP reported this confidence interval when the hypothesis testing procedure was conducted.

William Sealy Gosset (1876, 1937)

Gosset's interest in barley cultivation led him to speculate that design of experiments should aim, not only at improving the average yield, but also at breeding varieties whose yield was insensitive (robust) to variation in soil and climate.

Developed the *t*-test (1908)

Gosset was a friend of both Karl Pearson and R.A. Fisher, an achievement, for each had a monumental ego and a loathing for the other.

Gosset was a modest man who cut short an admirer with the comment that "Fisher would have discovered it all anyway."

'Student' in 1908

Computer Two-Sample *t*-Test Results

■ TABLE 2.2
Computer Output for the Two-Sample *t*-Test

```
Minitab
Two-sample T for Modified vs Unmodified

                 N        Mean        Std. Dev.     SE Mean
Modified        10      16.764            0.316        0.10
Unmodified      10      17.042            0.248       0.078

Difference = mu (Modified) − mu (Unmodified)
Estimate for difference: −0.278000
95% CI for difference: (−0.545073, −0.010927)
T-Test of difference = 0 (vs not = ): T-Value = −2.19
P-Value = 0.042  DF = 18
Both use Pooled Std. Dev. = 0.2843

JMP t-test

Unmodified-Modified

Assuming equal variances

Difference       0.278000  t Ratio     2.186876
Std Err Dif      0.127122  DF                18
Upper CL Dif     0.545073  Prob > |t|    0.0422
Lower CL Dif     0.010927  Prob > t      0.0211
Confidence           0.95  Prob < t      0.9789
```