

Probability and Statistics

Prepared by Dr. Alfaisal Hasan

STA 201
Probability and Statistics

Lecture NO 8



Probability and Statistics

Lecture NO 8

STA 201

Probability and Statistics

Prepared by
Dr. Alfaisal Hasan

www.midocean.ae

Chapter

Topics

- Chapter 1: Introduction To Statistics
- Chapter 2: Descriptive Statistics
- Chapter 3: Probability
- Chapter 4: Discrete Probability Distributions
- Chapter 5: Normal Probability Distributions
- Chapter 6: Confidence Intervals
- Chapter 7: Hypothesis Testing with One Sample
- **Chapter 8: Hypothesis Testing with Two Samples**
- Chapter 9: Correlation And Regression
- Chapter 10: Chi-Square Tests and The F -Distribution

الفصل الأول: مقدمة في الإحصاء

الفصل الثاني: الإحصاء الوصفي

الفصل الثالث: الاحتمالات

الفصل الرابع: التوزيعات الاحتمالية المنفصلة

الفصل الخامس: التوزيع الاحتمالي الطبيعي

الفصل السادس: فترات الثقة

الفصل السابع: اختبار الفرضيات لعينة واحدة

الفصل الثامن: اختبار الفرضيات لعينتين

الفصل التاسع: الارتباط والانحدار

الفصل العاشر: اختبارات كاي تربيع وتوزيع F

Hypothesis Testing with Two Sample

8.1 Testing the Difference Between Means (Large Independent Samples)

8.1 اختبار الفرق بين الوسائل (عينات مستقلة كبيرة)

8.2 Testing the Difference Between Means (Small Independent Samples)

8.2 اختبار الفرق بين الوسائل (عينات صغيرة مستقلة)

8.3 Testing the Difference Between Means (Dependent Samples)

8.3 اختبار الفرق بين الوسائل (العينات التابعة)

8.4 اختبار الفرق بين النسب

8.4 Testing the Difference Between Proportions

8.1 Testing the Difference Between Means (Large Independent Samples)

- Determine whether two samples are independent or dependent
تحديد ما إذا كانت عينتان مستقلتان أم معتمدتان
- Perform a two-sample z -test for the difference between two means μ_1 and μ_2 using large independent samples

Independent and Dependent Samples

Chapter

Independent Samples

The sample selected from one population is not related to the sample selected from the second population.

Dependent Samples (paired or matched samples)

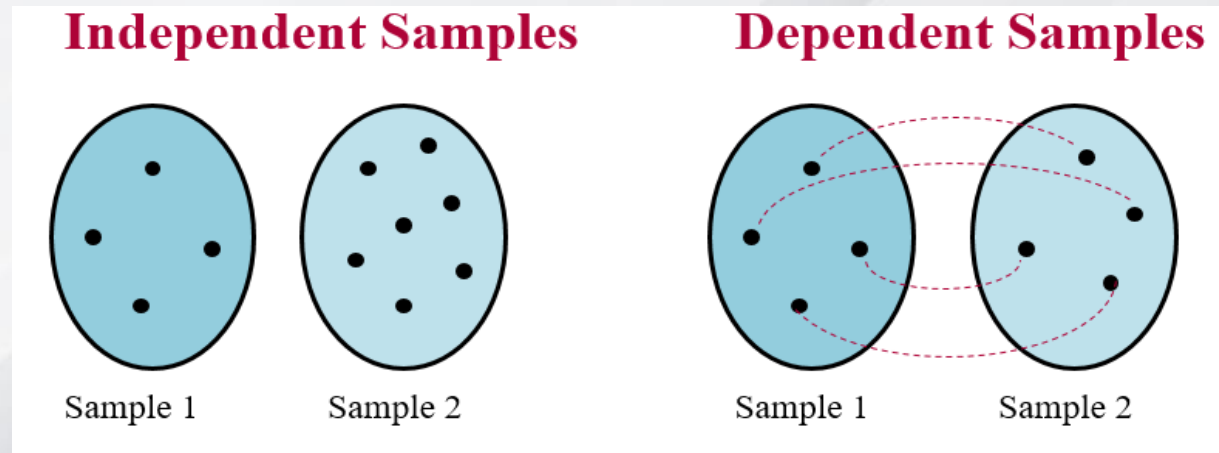
Each member of one sample corresponds to a member of the other sample.

عينات مستقلة

العينة المختارة من مجتمع واحد لا ترتبط بالعينة المختارة من المجتمع الثاني.

العينات التابعة (العينات المقترنة أو المتطابقة)

كل عضو في عينة واحدة يتوافق مع عضو في العينة الأخرى.



Example: Independent and Dependent Samples

Classify the pair of samples as independent or dependent.

صنف زوج العينات إلى مستقل أو تابع.

- Sample 1: Resting heart rates of 35 individuals before drinking coffee.
العينة 1: معدلات ضربات القلب لـ 35 فردًا قبل شرب القهوة.
- Sample 2: Resting heart rates of the same individuals after drinking two cups of coffee.
العينة 2: معدل ضربات القلب أثناء الراحة لنفس الأفراد بعد شرب فنجانين من القهوة.

Solution:

Dependent Samples (The samples can be paired with respect to each individual)
العينات التابعة (يمكن إقران العينات فيما يتعلق بكل فرد)

Example: Independent and Dependent Samples

Classify the pair of samples as independent or dependent.
صنف زوج العينات إلى مستقل أو تابع.

- Sample 1: Test scores for 35 statistics students.
العينة 1: درجات الاختبار لـ 35 طالبًا في مجال الإحصاء
- Sample 2: Test scores for 42 biology students who do not study statistics.
العينة 2: درجات الاختبار لـ 42 طالبًا من طلاب الأحياء الذين لا يدرسون الإحصاء.

Solution:

Independent Samples (Not possible to form a pairing between the members of the samples; the sample sizes are different, and the data represent scores for different individuals.)

العينات المستقلة (ليس من الممكن تشكيل اقتران بين أعضاء العينات؛ لأن أحجام العينات مختلفة، وتمثل البيانات درجات لأفراد مختلفين.)

Two Sample Hypothesis Test with Independent Samples

1. Null hypothesis H_0

A statistical hypothesis that usually states there is no difference between the parameters of two populations.

Always contains the symbol $\leq$, $=$, or $\geq$.

فرضية إحصائية تنص عادة على عدم وجود فرق بين معلمات مجموعتين من السكان.

2. Alternative hypothesis H_a

A statistical hypothesis that is true when H_0 is false.

Always contains the symbol $>$, $\neq$, or $<$.

فرضية إحصائية تكون صحيحة عندما الفرض الصفري يكون خاطئة.

Two Sample Hypothesis Test with Independent Samples

$$\begin{array}{lll} \left\{ \begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 \neq \mu_2 \end{array} \right. & \left\{ \begin{array}{l} H_0: \mu_1 \leq \mu_2 \\ H_a: \mu_1 > \mu_2 \end{array} \right. & \left\{ \begin{array}{l} H_0: \mu_1 \geq \mu_2 \\ H_a: \mu_1 < \mu_2 \end{array} \right. \end{array}$$

Regardless of which hypotheses you use, you always assume there is no difference between the population means, or $\mu_1 = \mu_2$.

Two Sample z-Test for the Difference Between Means Large Samples

Three conditions are necessary to perform a z-test for the difference between two population means μ_1 and μ_2 .

هناك ثلاثة شروط ضرورية لإجراء اختبار الفرق بين
وسطين سكانيين

1. The samples must be randomly selected.
2. The samples must be independent.
3. Each sample size must be at least 30, or, if not, each population must have a normal distribution with a known standard deviation.

يجب أن يتم اختيار العينات عشوائياً.
يجب أن تكون العينات مستقلة.
يجب أن يكون حجم كل عينة 30 على الأقل، وإذا لم يكن
الأمر كذلك، يجب أن يكون لكل مجتمع توزيع طبيعي مع
انحراف معياري معروف.

Steps for conducting Testing the Difference Between Means (Large Independent Samples)

1- State H_0 and H_a

2- identify α

3- Critical value

4- Test statistic

5- Take decision

Two Sample z-Test for the Difference Between Means

The **test statistic** is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

When the samples are large, you can use s_1 and s_2 in place of σ_1 and σ_2 . If the samples are not large, you can still use a two-sample z-test, provided the populations are normally distributed and the population standard deviations are known.

Example: Two-Sample z-Test for the Difference Between Means

A consumer education organization claims that there is a difference in the mean credit card debt of males and females in the United States. The results of a random survey of 200 individuals from each group are shown below. The two samples are independent. Do the results support the organization's claim? Use $\alpha = 0.05$.

تدعي منظمة تثقيف المستهلك أن هناك فرقاً في متوسط ديون بطاقات الائتمان بين الذكور والإناث في الولايات المتحدة. وفيما يلي نتائج المسح العشوائي الذي شمل 200 فرد من كل مجموعة. العينتان مستقلتان. هل النتائج تدعم ادعاء المنظمة؟
 $\alpha = 0.05$



Females (1)

$$\bar{x}_1 = \$2290$$

$$s_1 = \$750$$

$$n_1 = 200$$

Males (2)

$$\bar{x}_2 = \$2370$$

$$s_2 = \$800$$

$$n_2 = 200$$

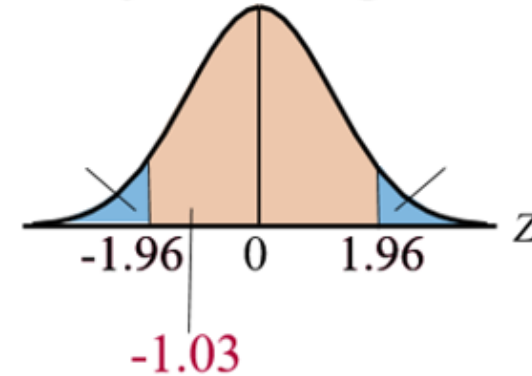


Solution: Two-Sample z-Test for the Difference Between Means

- $H_0: \mu_1 = \mu_2$
- $H_a: \mu_1 \neq \mu_2$
- $\alpha = 0.05$
- $n_1 = 200, n_2 = 200$
- critical value = -1.96 and 1.96
- **Test Statistic:**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$
$$z = \frac{(2290 - 2370) - 0}{\sqrt{\frac{750^2}{200} + \frac{800^2}{200}}} = -1.03$$

- **Rejection Region:**



- **Decision: Fail to Reject H_0**
At the 5% level of significance, there is not enough evidence to support the organization's claim that there is a difference in the mean credit card debt of males and females.

8.2 Testing the Difference Between Means (Small Independent Samples)

If samples of size less than 30 are taken from normally-distributed populations, a t -test may be used to test the difference between the population means μ_1 and μ_2 .

إذا تم أخذ عينات بحجم أقل من 30 من مجموعات سكانية موزعة بشكل طبيعي، فيمكن استخدامها لاختبار الفرق بين متوسطات السكان.

Three conditions are necessary to use a t -test for small independent samples:

1. The samples must be randomly selected.
2. The samples must be independent.
3. Each population must have a normal distribution.

هناك ثلاثة شروط ضرورية لاستخدام للعينات المستقلة الصغيرة:
يجب أن يتم اختيار العينات عشوائياً.
يجب أن تكون العينات مستقلة.
يجب أن يكون لكل مجتمع التوزيع الطبيعي.

Testing the Difference Between Means (Small Independent Samples)

The **standardized test statistic** is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

The standard error and the degrees of freedom of the sampling distribution depend on whether the population variances are equal.

Testing the Difference Between Means (Small Independent Samples)

- **Variances are equal**

- The standard error for the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- d.f. = $n_1 + n_2 - 2$

- **Variances are not equal**

- the standard error is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- d.f = smaller of $n_1 - 1$ or $n_2 - 1$

Test Statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$


- **Variances are equal**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- d.f. = $n_1 + n_2 - 2$

- **Variances are not equal**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- d.f = smaller of $n_1 - 1$ or $n_2 - 1$

Example: Two-Sample t-Test for the Difference Between Means

A manufacturer claims that the calling range (in feet) of its 2.4-GHz cordless telephone is greater than that of its leading competitor. You perform a study using 14 randomly selected phones from the manufacturer and 16 randomly selected similar phones from its competitor. The results are shown below. At $\alpha = 0.05$, can you support the manufacturer's claim? Assume the populations are normally distributed and the population variances are equal.

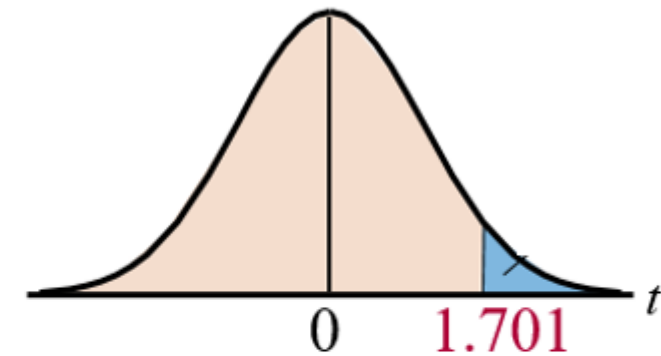
تدعي إحدى الشركات المصنعة أن نطاق الاتصال (بالقدم) للهاتف اللاسلكي الذي يعمل بتردد 2.4 جيجا هرتز أكبر من نطاق الاتصال الخاص بمنافسها الرئيسي. يمكنك إجراء دراسة باستخدام 14 هاتفًا تم اختيارها عشوائيًا من الشركة المصنعة و16 هاتفًا مشابهًا تم اختيارها عشوائيًا من منافسها. النتائج مبينة أدناه. هل يمكنك دعم مطالبة الشركة المصنعة؟ افترض أن السكان موزعون بشكل طبيعي وأن تباينات السكان متساوية.

Manufacturer (1)	Competition (2)
$\bar{x}_1 = 1275 \text{ ft}$	$\bar{x}_2 = 1250 \text{ ft}$
$s_1 = 45 \text{ ft}$	$s_2 = 30 \text{ ft}$
$n_1 = 14$	$n_2 = 16$

Solution: Two-Sample t-Test for the Difference Between Means

- $H_0: \mu_1 \leq \mu_2$
- $H_a: \mu_1 > \mu_2$
- $\alpha = 0.05$
- $\text{d.f.} = 14 + 16 - 2 = 28$
- critical value = 1.701
- Test Statistic: ?

- Rejection Region: ?



- Decision: ?

T- distribution table

Chapter

	Level of confidence, c	0.50	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		.816	1.886	2.920	4.303	6.965	9.925
3		.765	1.638	2.353	3.182	4.541	5.841
4		.741	1.533	2.132	2.776	3.747	4.604
5		.727	1.476	2.015	2.571	3.365	4.032
6		.718	1.440	1.943	2.447	3.143	3.707
7		.711	1.415	1.895	2.365	2.998	3.499
8		.706	1.397	1.860	2.306	2.896	3.355
9		.703	1.383	1.833	2.262	2.821	3.250
10		.700	1.372	1.812	2.228	2.764	3.169
11		.697	1.363	1.796	2.201	2.718	3.106
12		.695	1.356	1.782	2.179	2.681	3.055
13		.694	1.350	1.771	2.160	2.650	3.012
14		.692	1.345	1.761	2.145	2.624	2.977
15		.691	1.341	1.753	2.131	2.602	2.947
16		.690	1.337	1.746	2.120	2.583	2.921
17		.689	1.333	1.740	2.110	2.567	2.898
18		.688	1.330	1.734	2.101	2.552	2.878
19		.688	1.328	1.729	2.093	2.539	2.861
20		.687	1.325	1.725	2.086	2.528	2.845
21		.686	1.323	1.721	2.080	2.518	2.831
22		.686	1.321	1.717	2.074	2.508	2.819
23		.685	1.319	1.714	2.069	2.500	2.807
24		.685	1.318	1.711	2.064	2.492	2.797
25		.684	1.316	1.708	2.060	2.485	2.787
26		.684	1.315	1.706	2.056	2.479	2.779
27		.684	1.314	1.703	2.052	2.473	2.771
28		.683	1.313	1.701	2.048	2.467	2.763
29		.683	1.311	1.699	2.045	2.462	2.756
∞		.674	1.282	1.645	1.960	2.326	2.576

Solution: Obtaining Test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

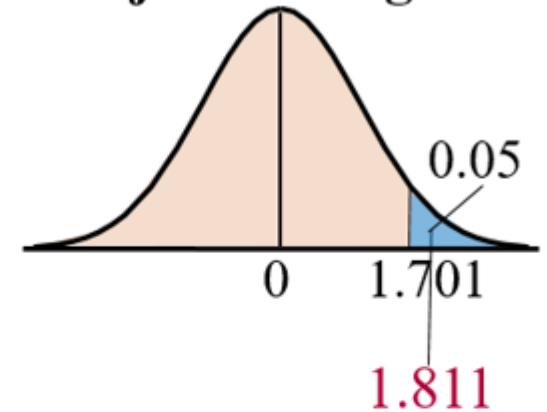
$$t = \frac{(1275 - 1250) - 0}{\sqrt{\frac{(14 - 1)(45)^2 + (16 - 1)(30)^2}{14 + 16 - 2}} \cdot \sqrt{\frac{1}{14} + \frac{1}{16}}}$$

$$t = \frac{(1275 - 1250) - 0}{13.8018} \approx 1.811$$

Solution: Two-Sample t-Test for the Difference Between Means

- $H_0: \mu_1 \leq \mu_2$
- $H_a: \mu_1 > \mu_2$
- $\alpha = 0.05$
- $\text{d.f.} = 14 + 16 - 2 = 28$
- critical value = 1.701
- **Test Statistic:**
 $t = 1.811$

- **Rejection Region:**



- **Decision: Reject H_0**
At the 5% level of significance, there is enough evidence to support the manufacturer's claim that its phone has a greater calling range than its competitors.

Example: Two-Sample t-Test for the Difference Between Means

The braking distances of 8 Volkswagen GTIs and 10 Ford Focuses were tested when traveling at 60 miles per hour on dry pavement. The results are shown below. Can you conclude that there is a difference in the mean braking distances of the two types of cars? Use $\alpha = 0.01$. Assume the populations are normally distributed and the population variances are not equal. *(Adapted from Consumer Reports)*

تم اختبار مسافات الكبح لـ 8 سيارات فولكس فاجن جي تي آي و 10 سيارات فورد فوكس عند السفر بسرعة 60 ميلاً في الساعة على الرصيف الجاف. النتائج مبينة أدناه. هل يمكنك استنتاج وجود اختلاف في متوسط مسافات الفرملة بين نوعي السيارات؟ افترض أن السكان موزعون بشكل طبيعي وأن تباينات السكان ليست متساوية. (مقتبس من تقارير المستهلك)

GTI (1)	Focus (2)
$\bar{x}_1 = 134 \text{ ft}$	$\bar{x}_2 = 143 \text{ ft}$
$s_1 = 6.9 \text{ ft}$	$s_2 = 2.6 \text{ ft}$
$n_1 = 8$	$n_2 = 10$

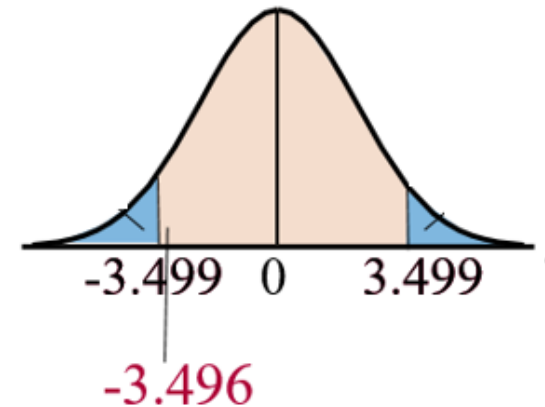
Solution: Two-Sample t-Test for

- $H_0: \mu_1 = \mu_2$
- $H_a: \mu_1 \neq \mu_2$
- $\alpha = 0.01$
- $\text{d.f.} = 8 - 1 = 7$
- critical value = -3.499 and 3.499

- **Test Statistic:**

$$t = \frac{(134 - 143) - 0}{\sqrt{\frac{6.9^2}{8} + \frac{2.6^2}{10}}} = -3.496$$

- **Rejection Region:**



- **Decision: Fail to Reject H_0**

At the 1% level of significance, there is not enough evidence to conclude that the mean braking distances of the cars are different.

T- distribution table

Chapter

	Level of confidence, c	0.50	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		.816	1.886	2.920	4.303	6.965	9.925
3		.765	1.638	2.353	3.182	4.541	5.841
4		.741	1.533	2.132	2.776	3.747	4.604
5		.727	1.476	2.015	2.571	3.365	4.032
6		.718	1.440	1.943	2.447	3.143	3.707
7		.711	1.415	1.895	2.365	2.998	3.499
8		.706	1.397	1.860	2.306	2.896	3.355
9		.703	1.383	1.833	2.262	2.821	3.250
10		.700	1.372	1.812	2.228	2.764	3.169
11		.697	1.363	1.796	2.201	2.718	3.106
12		.695	1.356	1.782	2.179	2.681	3.055
13		.694	1.350	1.771	2.160	2.650	3.012
14		.692	1.345	1.761	2.145	2.624	2.977
15		.691	1.341	1.753	2.131	2.602	2.947
16		.690	1.337	1.746	2.120	2.583	2.921
17		.689	1.333	1.740	2.110	2.567	2.898
18		.688	1.330	1.734	2.101	2.552	2.878
19		.688	1.328	1.729	2.093	2.539	2.861
20		.687	1.325	1.725	2.086	2.528	2.845
21		.686	1.323	1.721	2.080	2.518	2.831
22		.686	1.321	1.717	2.074	2.508	2.819
23		.685	1.319	1.714	2.069	2.500	2.807
24		.685	1.318	1.711	2.064	2.492	2.797
25		.684	1.316	1.708	2.060	2.485	2.787
26		.684	1.315	1.706	2.056	2.479	2.779
27		.684	1.314	1.703	2.052	2.473	2.771
28		.683	1.313	1.701	2.048	2.467	2.763
29		.683	1.311	1.699	2.045	2.462	2.756
∞		.674	1.282	1.645	1.960	2.326	2.576

Section 8.3: Testing the Difference Between Means (Dependent Samples)

- To perform a two-sample hypothesis test with dependent samples, the difference between each data pair is first found:
 - $d = x_1 - x_2$ Difference between entries for a data pair
- The test statistic is the mean $\bar{d}$ of these differences.
 - $\bar{d} = \frac{\sum d}{n}$ Mean of the differences between paired data entries in the dependent samples
- The standard deviation of the differences between the paired data entries in the dependent samples

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}} = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n - 1}}$$

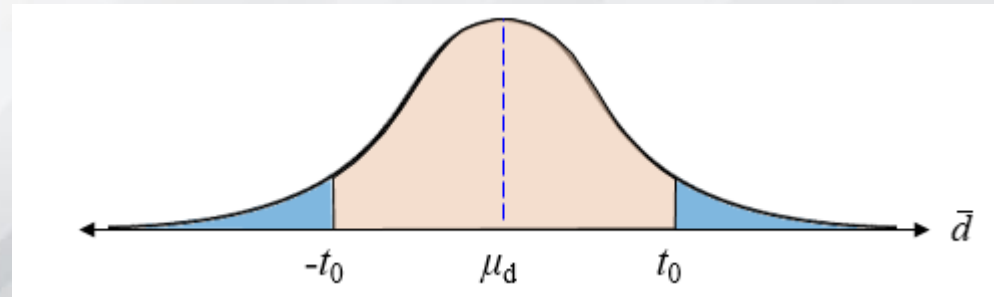
Testing the Difference Between Means (Dependent Samples)

Three conditions are required to conduct the test.

1. The samples must be randomly selected.
2. The samples must be dependent (paired).
3. Both populations must be normally distributed.

If these conditions are met, then the sampling distribution for $\bar{d}$ is approximated by a t -distribution with $n - 1$ degrees of freedom, where n is the number of data pairs.

هناك ثلاثة شروط مطلوبة لإجراء الاختبار.
يجب أن يتم اختيار العينات عشوائياً.
يجب أن تكون العينات تابعة (مقترنة).
يجب أن يتم توزيع كلا المجموعتين بشكل طبيعي.



Testing the Difference Between Means (Dependent Samples)

- The **test statistic** is $\bar{d} = \frac{\sum d}{n}$

- The **standardized test statistic** is

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

- The **degrees of freedom** are

$$\text{d.f.} = n - 1$$

Null Hypothesis and Alternative Hypothesis

First Case:

$$H_0: \mu_d =$$

$$H_a: \mu_d \neq$$

Second Case:

$$H_0: \mu_d \leq$$

$$H_a: \mu_d >$$

Third Case:

$$H_0: \mu_d \geq$$

$$H_a: \mu_d <$$

Example: Testing the Difference Between Means (Dependent Samples)

A golf club manufacturer claims that old scores are greater than new scores even if using the manufacturer's newly designed golf clubs. Eight golfers are randomly selected, and each is asked to give his or her most recent score. After using the new clubs for one month, the golfers are again asked to give their most recent score. The scores for each golfer are shown in the table. Assuming the golf scores are normally distributed, is there enough evidence to support the manufacturer's claim at $\alpha = 0.10$?

يدعي أحد مصنعي مضارب الجولف أن النتائج القديمة أكبر من النتائج الجديدة حتى في حالة استخدام مضارب الجولف المصممة حديثاً من قبل الشركة المصنعة. يتم اختيار ثمانية من لاعبي الغولف بشكل عشوائي، ويطلب من كل منهم تقديم أحدث نتيجة له. بعد استخدام المضارب الجديدة لمدة شهر واحد، يُطلب من لاعبي الجولف مرة أخرى تقديم أحدث نتائجهم. يتم عرض النتائج لكل لاعب غولف في الجدول. على افتراض أن نتائج لعبة الجولف يتم توزيعها بشكل طبيعي، فهل هناك أدلة كافية لدعم ادعاء الشركة المصنعة
 $\alpha = 0.10$

Golfer	1	2	3	4	5	6	7	8
Score (old)	89	84	96	82	74	92	85	91
Score (new)	83	83	92	84	76	91	80	91

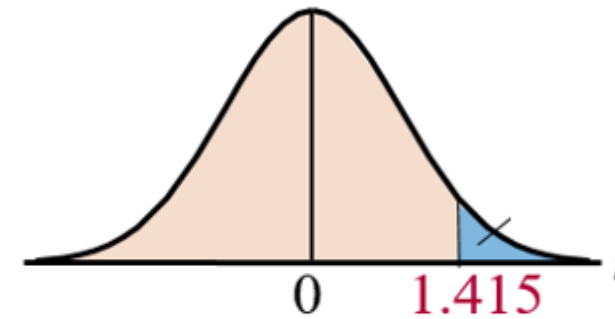
Solution: Testing the Difference Between Means (Dependent Samples)

- $H_0: \mu_d \leq 0$
- $H_a: \mu_d > 0$
- $\alpha = 0.10$
- $d.f. = 8 - 1 = 7$

• critical value = 1.415

• Test Statistic: ?

- Rejection Region:



• Decision: ?

T- distribution table

Chapter

	Level of confidence, c	0.50	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		.816	1.886	2.920	4.303	6.965	9.925
3		.765	1.638	2.353	3.182	4.541	5.841
4		.741	1.533	2.132	2.776	3.747	4.604
5		.727	1.476	2.015	2.571	3.365	4.032
6		.718	1.440	1.943	2.447	3.143	3.707
7		.711	1.415	1.895	2.365	2.998	3.499
8		.706	1.397	1.860	2.306	2.896	3.355
9		.703	1.383	1.833	2.262	2.821	3.250
10		.700	1.372	1.812	2.228	2.764	3.169
11		.697	1.363	1.796	2.201	2.718	3.106
12		.695	1.356	1.782	2.179	2.681	3.055
13		.694	1.350	1.771	2.160	2.650	3.012
14		.692	1.345	1.761	2.145	2.624	2.977
15		.691	1.341	1.753	2.131	2.602	2.947
16		.690	1.337	1.746	2.120	2.583	2.921
17		.689	1.333	1.740	2.110	2.567	2.898
18		.688	1.330	1.734	2.101	2.552	2.878
19		.688	1.328	1.729	2.093	2.539	2.861
20		.687	1.325	1.725	2.086	2.528	2.845
21		.686	1.323	1.721	2.080	2.518	2.831
22		.686	1.321	1.717	2.074	2.508	2.819
23		.685	1.319	1.714	2.069	2.500	2.807
24		.685	1.318	1.711	2.064	2.492	2.797
25		.684	1.316	1.708	2.060	2.485	2.787
26		.684	1.315	1.706	2.056	2.479	2.779
27		.684	1.314	1.703	2.052	2.473	2.771
28		.683	1.313	1.701	2.048	2.467	2.763
29		.683	1.311	1.699	2.045	2.462	2.756
∞		.674	1.282	1.645	1.960	2.326	2.576

Solution: Obtaining test statistic

$d = (\text{old score}) - (\text{new score})$

Old	New	d	d^2
89	83	6	36
84	83	1	1
96	92	4	16
82	84	-2	4
74	76	-2	4
92	91	1	1
85	80	5	25
91	91	0	0
		$\Sigma = 13$	$\Sigma = 87$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{13}{8} = 1.625$$

$$\begin{aligned} s_d &= \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}} \\ &= \sqrt{\frac{87 - \frac{(13)^2}{8}}{8-1}} \\ &\approx 3.0677 \end{aligned}$$

- **Test Statistic:**

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \approx \frac{1.625 - 0}{3.0677 / \sqrt{8}} \approx 1.498$$

Solution: Testing the Difference Between Means (Dependent Samples)

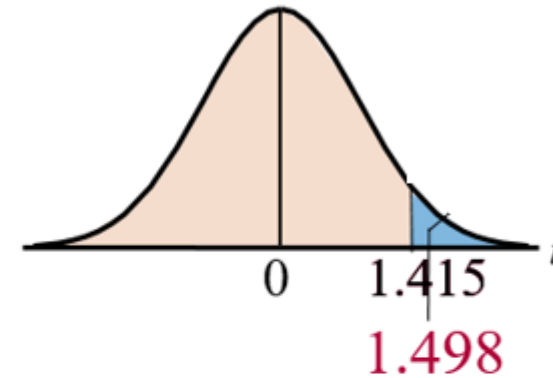
- $H_0: \mu_d \leq 0$
- $H_a: \mu_d > 0$
- $\alpha = 0.10$
- $\text{d.f.} = 8 - 1 = 7$

- critical value = 1.415

- Test Statistic:

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \approx \frac{1.625 - 0}{3.0677 / \sqrt{8}} \approx 1.498$$

- Rejection Region:



- Decision: **Reject H_0**

At the 10% level of significance, the results of this test indicate that after the golfers used the new clubs, their scores were significantly lower.

Section 8.4: Testing the Difference Between Proportions

Used to test the difference between two population proportions, p_1 and p_2 .

Three conditions are required to conduct the test:

1. The samples must be randomly selected.
2. The samples must be independent.
3. The samples must be large enough to use a normal sampling distribution. That is,
 $n_1p_1 \geq 5$, $n_1q_1 \geq 5$, $n_2p_2 \geq 5$, and $n_2q_2 \geq 5$.

Testing the Difference Between Proportions

- If these conditions are met, then the sampling distribution for $\hat{p}_1 - \hat{p}_2$ is a normal distribution
- Mean: $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$
- A weighted estimate of p_1 and p_2 can be found by using $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$, where $x_1 = n_1 \hat{p}_1$ and $x_2 = n_2 \hat{p}_2$

- Standard error:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Testing the Difference Between Proportions

- The **test statistic** is $\hat{p}_1 - \hat{p}_2$
- The **standardized test statistic** is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \text{ and } \bar{q} = 1 - \bar{p}$$

Note: $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, and $n_2\bar{q}$ must be at least 5.

Example: Testing the Difference Between Proportions

In a study of 200 randomly selected adult female and 250 randomly selected adult male Internet users, 30% of the females and 38% of the males said that they plan to shop online at least once during the next month. At $\alpha = 0.10$ test the claim that there is a difference between the proportion of female and the proportion of male Internet users who plan to shop online.

في دراسة أجريت على 200 أنثى بالغة تم اختيارها عشوائيًا و250 من مستخدمي الإنترنت البالغين الذين تم اختيارهم عشوائيًا، قال 30% من الإناث و38% من الذكور أنهم يخططون للتسوق عبر الإنترنت مرة واحدة على الأقل خلال الشهر المقبل. اختبار الادعاء بوجود فرق بين نسبة الإناث ونسبة مستخدمي الإنترنت من الذكور الذين يخططون للتسوق عبر الإنترنت.

Solution:

1 = Females

2 = Males

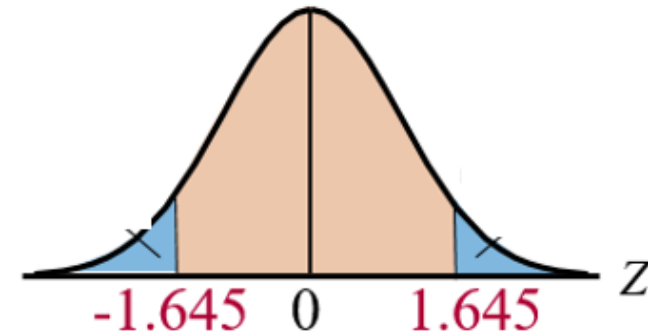
Solution: Testing the Difference Between Proportions

- $H_0: p_1 = p_2$
- $H_a: p_1 \neq p_2$
- $\alpha = 0.10$
- $n_1 = 200, n_2 = 250$

- critical value = -1.645 and 1.645

- Test Statistic: ?

- Rejection Region:



- Decision: ?

Solution: Obtaining test statistic

$$x_1 = n_1 \hat{p}_1 = 60 \quad x_2 = n_2 \hat{p}_2 = 95$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{60 + 95}{200 + 250} \approx 0.3444$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.3444 = 0.6556$$

Note:

$$n_1 \bar{p} = 200(0.3444) \geq 5 \quad n_1 \bar{q} = 200(0.6556) \geq 5$$

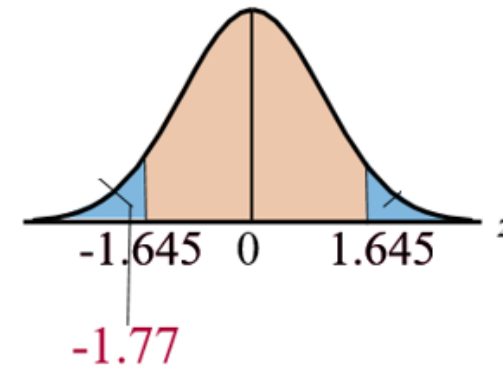
$$n_2 \bar{p} = 250(0.3444) \geq 5 \quad n_2 \bar{q} = 250(0.6556) \geq 5$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \approx \frac{(0.30 - 0.38) - (0)}{\sqrt{(0.3444) \cdot (0.6556) \cdot \left(\frac{1}{200} + \frac{1}{250} \right)}} \\ \approx -1.77$$

Solution: Testing the Difference Between Proportions

- $H_0: p_1 = p_2$
- $H_a: p_1 \neq p_2$
- $\alpha = 0.10$
- $n_1 = 200, n_2 = 250$
- critical value = -1.645 and 1.645
- Test Statistic:
 $z = -1.77$

- Rejection Region:



- Decision: **Reject H_0**

At the 10% level of significance, there is enough evidence to conclude that there is a difference between the proportion of female and the proportion of male Internet users who plan to shop online.

Example: Testing the Difference Between Proportions

A medical research team conducted a study to test the effect of a cholesterol reducing medication. At the end of the study, the researchers found that of the 4700 randomly selected subjects who took the medication, 301 died of heart disease. Of the 4300 randomly selected subjects who took a placebo, 357 died of heart disease. At $\alpha = 0.01$ can you conclude that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo? *(Adapted from New England Journal of Medicine)*

أجرى فريق بحث طبي دراسة لاختبار تأثير دواء مخفض للكوليسترول. وفي نهاية الدراسة، وجد الباحثون أنه من بين 4700 شخص تم اختيارهم عشوائيًا والذين تناولوا الدواء، توفي 301 منهم بسبب أمراض القلب. ومن بين 4300 شخص تم اختيارهم عشوائيًا والذين تناولوا الدواء الوهمي، توفي 357 منهم بسبب أمراض القلب. هل يمكنك أن تستنتج أن معدل الوفيات بسبب أمراض القلب أقل لدى أولئك الذين تناولوا الدواء مقارنة بمن تناولوا الدواء الوهمي؟ (مقتبس من مجلة نيو إنغلاند الطبية)

Solution:

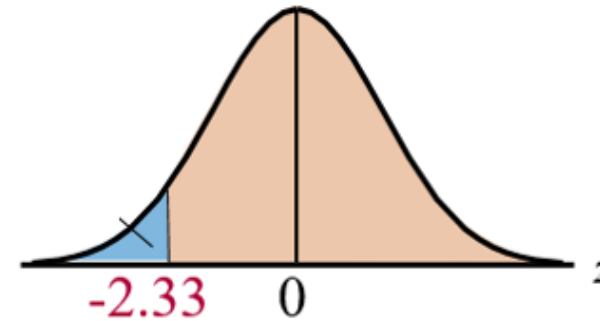
1 = Medication

2 = Placebo

Solution: Testing the Difference Between Proportions

- $H_0: p_1 \geq p_2$
- $H_a: p_1 < p_2$
- $\alpha = 0.01$
- $n_1 = 4700, n_2 = 4300$
- critical value = -2.33
- Test Statistic: ?

- Rejection Region: ?



- Decision: ?

Solution: Obtaining test statistic

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{301}{4700} = 0.064 \quad \hat{p}_2 = \frac{x_2}{n_2} = \frac{357}{4300} = 0.083$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{301 + 357}{4700 + 4300} \approx 0.0731$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.0731 = 0.9269$$

Note:

$$n_1\bar{p} = 4700(0.0731) \geq 5 \quad n_1\bar{q} = 4700(0.9269) \geq 5$$

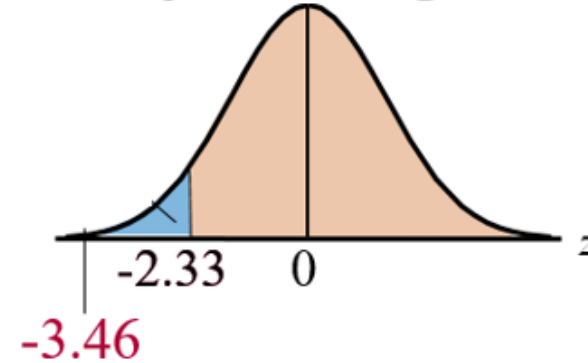
$$n_2\bar{p} = 4300(0.0731) \geq 5 \quad n_2\bar{q} = 4300(0.9269) \geq 5$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \frac{(0.064 - 0.083) - (0)}{\sqrt{(0.0731) \cdot (0.9269) \cdot \left(\frac{1}{4700} + \frac{1}{4300}\right)}} \\ \approx -3.46$$

Solution: Testing the Difference Between Proportions

- $H_0: p_1 \geq p_2$
- $H_a: p_1 < p_2$
- $\alpha = 0.01$
- $n_1 = 4700, n_2 = 4300$
- critical value = -2.33
- **Test Statistic:**
 $z = -3.46$

- **Rejection Region:**



- **Decision: Reject H_0**
At the 1% level of significance, there is enough evidence to conclude that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo.

Chapter Summary

Chapter

8.1 Testing the Difference Between Means (Large Independent Samples)

8.2 Testing the Difference Between Means (Small Independent Samples)

8.3 Testing the Difference Between Means (Dependent Samples)

8.4 Testing the Difference Between Proportions

Un Solved Examples

Chapter

No need for the degrees of freedom When obtaining the critical value using
t test

لا يوجد احتياج لدرجات الحرية عند الحصول على القيمة الحرجة باستخدام اختبار t

Un Solved Examples

Chapter

If $H_0 : P_1 = P_2, \alpha=0.1, n_1 = 200$ and $n_2 = 250$ critical value -1.6 and 1.6 and test statistics $z = -1.9$, then we fail to reject H_0

إذا كان $H_0 : P_1 = P_2, \alpha=0.1, n_1 = 200$ and $n_2 = 250$ والقيمة الحرجة -1.6 and 1.6 والقيمة الناتجة من الاختبار $z = -1.9$ فإننا نرفض الفرض الصفري



Thank You