Statistikë e aplikuar

UNIVERSITETI I EJL
ЈИЕ УНИВЕРЗИТЕТ
SEE UNIVERSITY

Përfundimet statistikore mbështetur në dy mostra

Faton Berisha

Kapitulli 9

Përfundimet statistikore mbështetur në dy mostra

Përfundimet statistikore mbështetur në dy mostra

9.1 Krahasimi i dy mesatareve të popullimit duke shfrytëzuar mostra të mëdha të pavarura

9.2 Krahasimi i dy mesatareve të popullimit duke shfrytëzuar mostra të vogla të pavarura

9.3 Krahasimi i dy variancave të popullimit duke shfrytëzuar mostra të pavarura

9.4 Eksperimentet e diferencës së çiftëzuar

9.5 Krahasimi i dy proporcioneve të popullimit duke shfrytëzuar mostra të mëdha të pavarura

9.1 Shpërndarja e zgjedhjeve e $\bar{x}_1 - \bar{x}_2$

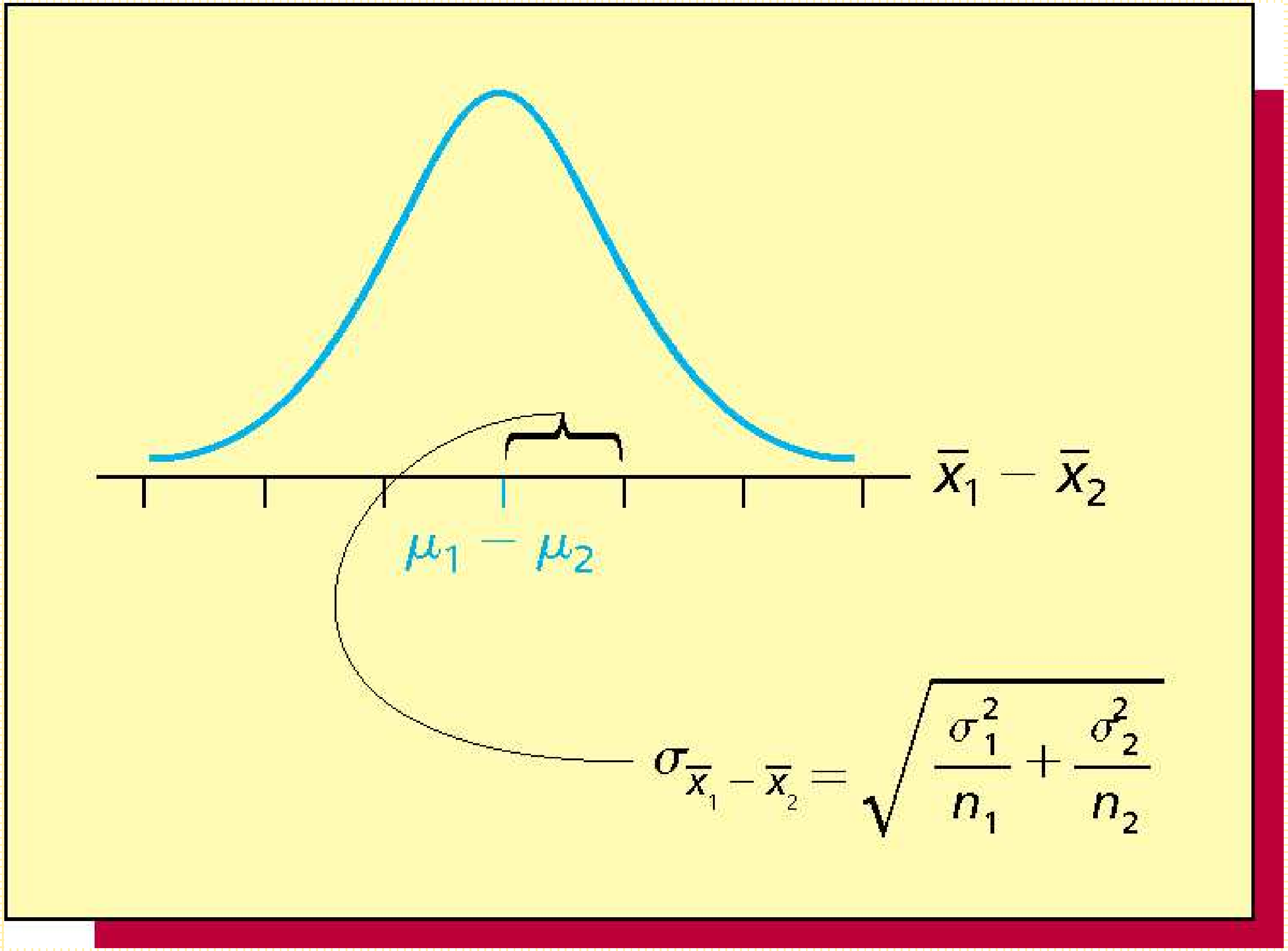
Në qoftë se merrem mostra të rastësishme të pavarura nga dy popullime, atëherë shpërndarja e zgjedhjeve e diferencës së mesatareve të mostrave $\bar{x}_1 - \bar{x}_2$ është

Normale, në qoftë se secili nga popullimet e zgjedhura është normal dhe përafërsisht normal në qoftë se madhësitë n_1 dhe n_2 janë të mëdha

Ka mesatare: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Ka devijim standard: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Shpërndarja e zgjedhjeve e $\bar{x}_1 - \bar{x}_2$ (Vazhdim)



Intervlai i besueshmërisë për mostër të madhe, diferenca e mesatareve

Në qoftë se dy mostrat e pavarura janë nga popullime që janë normale ose secilia nga madhësitë e mostrave është e madhe, **100(1 - α)% intervali i besueshërisë për μ_1 - μ_2** është

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Në qoftë se σ_1 dhe σ_2 janë të panjohura dhe secila nga madhësitë e mostrave është e madhe (n_1, $n_2 \geq 30$), vlerësojmë devijimet standarde të mostrave me s_1 dhe s_2, dhe **100(1 - α)% intervali i besueshmërisë pëe μ_1 - μ_2** është

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Testet e mostrave të mëdha mbi diferencën e mesatareve

Në qoftë se populimet e zgjedhura janë normale ose të dyja mostrat janë të mëdha, mund të **hedhim poshtë H_0: $\mu_1 - \mu_2 = D_0$ me nivel rëndësie** α atëherë dhe vetëm atëherë kur vlen vetia përkatëse pikësore ose, në mënyrë ekuivalente, në qoftë se p-vlera përkatëse është më e vogël se α.

Alternativa	**Hedh H_0 në qoftë se:**	**p-vlera**
$H_a: \mu_1 - \mu_2 > D_0$	$z > z_\alpha$	Syprina nën lakoren std. norm. në të djathtë të z
$H_a: \mu_1 - \mu_2 < D_0$	$z < -z_\alpha$	Syprina nën lakoren std. norm. në të majtë të z
$H_a: \mu_1 - \mu_2 \neq D_0$	$\lvert z \rvert > z_{\alpha/2}$, dmth. $z > z_{\alpha/2}$ ose $z < -z_{\alpha/2}$	Dyfish syprina nën lakoren std. norm. në të djathtë të $\lvert z$

Statistika e testit

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Nëse se varianca e populimit e panjohur, dhe madhësitë e mostrave të mëdha, zëvendëso variancat e mostrave.

Shembull: Intervali dhe testi për mostra të mëdha

Kohët e pritjes në bankë, Sistemi aktual kundër të riut

$$n_1 = 100, \bar{x}_1 = 8.79, s_1^2 = 4.7089, \quad n_2 = 100, \bar{x}_2 = 5.14, s_2^2 = 1.9044$$

95% intervali i besueshmërisë për μ_1 - μ_2

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (8.79 - 5.14) \pm 1.96\sqrt{\frac{4.7089}{100} + \frac{1.9044}{100}}$$

$$= 3.65 \pm 0.5040 = [3.15, 4.15]$$

Testi H_0: μ_1 - $\mu_2 \leq 3$ kundër H_a: μ_1 - $\mu_2 > 3$, $\alpha = 0.05$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(8.79 - 5.14) - 3}{\sqrt{\frac{4.7089}{100} + \frac{1.9044}{100}}} = 2.53 > 1.645 = z_{.05}$$

$$p - \text{vlera} = P(z > 2.53) = (0.5 - 0.4943) = 0.0057$$

9.2 Intervali i besueshmërisë për mostra të vogla, Diferenca e mesatareve me varianca të barabarta

Në qoftë se dy mostrat e pavarura janë nga popullime normale me varianca të barabarta, **100(1 - α)% intervali i besueshmërisë për μ_1 - μ_2** është

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

ku s_p^2 është varianca e përbashkët

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

dhe $t_{\alpha/2}$ mbështetet në $(n_1 + n_2 - 2)$ shkallë lirie.

Testet e mostrave të vogla mbi diferencën e mesatareve për varianca të barabarta

Në qoftë se dy mostrat e pavarura janë nga popullime normale me varianca të barabarta, mund të **hedhim poshtë H_0: $\mu_1 - \mu_2 = D_0$ me nivel rëndësie** α atëherë dhe vetëm atëherë kur vlen vetia përkatëse pikësore ose, në mënyrë ekuivalente, në qoftë se p-vlera përkatëse është më e vogël se α.

Alternativa	**Hedh H_0 në qoftë se:**	**p-vlera**
$H_a: \mu_1 - \mu_2 > D_0$	$t > t_\alpha$	Syprina nën lakoren std. norm. në të djathtë të t
$H_a: \mu_1 - \mu_2 < D_0$	$t < -t_\alpha$	Syprina nën lakoren std. norm. në të majtë të t
$H_a: \mu_1 - \mu_2 \neq D_0$	$\lvert t\rvert > t_{\alpha/2}$, that is $t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$	Dyfish syprina nën lakoren std. norm. në të djathtë të $\lvert t\rvert$

Statistika e testit

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Varianca e përbashkët

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

t_α, $t_{\alpha/2}$ and p-vlerat mbështeten në $(n_1 + n_2 - 2)$ df

Shembull: Testi i diferencës së mesatareve për mostra të vogla

Rasit i katalizatorit, Diferenca e kthimeve mesatare për orë?

$$\mathrm{n}_1 = 5,\ \bar{x}_1 = 811.0,\ s_1^2 = 386, \quad \mathrm{n}_2 = 5,\ \bar{x}_2 = 750.2,\ s_2^2 = 484.2$$

Testo H_0: $\mu_1 - \mu_2 = 0$ kundrejt H_a: $\mu_1 - \mu_2 \neq 0$, $\alpha = 0.01$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} = \frac{(5-1)386 + (5-1)484.2}{(5+5-2)} = 435.1$$

$$\mathrm{t} = \frac{(\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(811 - 750.2) - 0}{\sqrt{435.1\left(\frac{1}{5} + \frac{1}{5}\right)}} = 4.6087 > 3.355 = t_{.005}$$

$$\mathrm{p-vlera} = \mathrm{P(t > 4.6087)} = 0.0017$$

Intervalet dhe testet e mostrave të vogla mbi ndryshimet e mesatareve kur variancat nuk janë të barabarta

Në qoftë se të dyja popullimet e zgjedhura janë normale, por madhësitë e mostrave dhe variancat ndryshojnë bindshëm, vlerësimi dhe testimi për mostra të vogla mund të mbështetet në procedurën vijuese “me varianca jo të barabarta”.

Intervali i besueshmërisë

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Statistika e testit

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Si për intervalin ashtu dhe për testin, **shkallët e lirisë** janë të barabarta me

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

9.3 Krahasimi i variancave të dy popullimeve me anë të mostrave të pavarura (Njëanshëm)

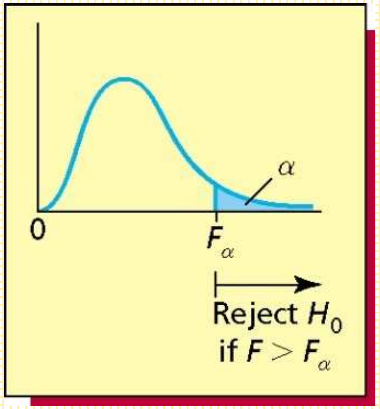
Në qoftë se të dyja popullimet e zgjedhura janë normale

Testi i $H_0: \sigma_1^2 \leq \sigma_2^2$ **kundrejt**
$H_a: \sigma_1^2 > \sigma_2^2$

Statistika e testit $F = \dfrac{s_1^2}{s_2^2}$

Hedh poshtë H_0 në qoftë se
$F > F_\alpha$ ose
p-vlera $< \alpha$

F_α mbështetet në
$(n_1 - 1)$ dhe $(n_2 - 1)$ df



Test of $H_0: \sigma_1^2 \geq \sigma_2^2$ **kundrejt**
$H_a: \sigma_1^2 < \sigma_2^2$

Statistika e testit $F = \dfrac{s_2^2}{s_1^2}$

Hedh poshtë H_0 në qoftë se
$F > F_\alpha$ ose
p-vlera $< \alpha$

F_α mbështetet në
$(n_2 - 1)$ dhe $(n_1 - 1)$ df

Krahasimi i variancave të dy popullimeve me anë të mostrave të pavarura (Dyanshëm)

Në qoftë se të dyja popullimet e zgjedhura janë normale

Testimi i H_0: $\sigma_1^2 = \sigma_2^2$ kundrejt H_a: $\sigma_1^2 \neq \sigma_2^2$

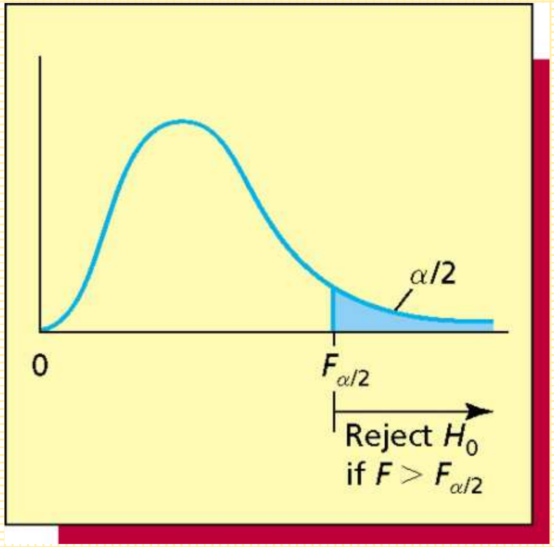
Statistika e tesit $$F = \frac{\text{më e madhja nga } s_1^2 \text{ dhe } s_2^2}{\text{më e vogla nga } s_1^2 \text{ dhe } s_2^2}$$

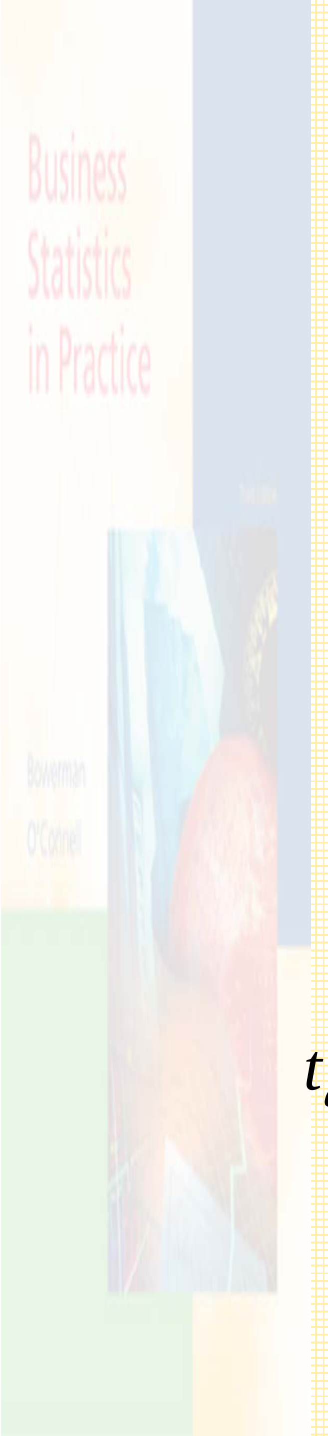
df_1 = {madhësia e mostrës me variancë më të madhe} – 1
df_2 = {madhësia e mostrës me variancë më vogël} – 1

Hedh poshtë H_0 në favor të H_a në qoftë se:

$F > F_{\alpha/2}$ ose

p-vlera $< \alpha$





9.4 Intervali i diferencës së çiftëzuar për diferencën e mesatareve

Në qoftë se popullimi i zgjdhur i diferencave është me shpërndarje normale me mesatare μ_d, atëherë **(1–α)100% intervali i besueshmërisë për $\mu_d = \mu_1 - \mu_2$** është

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$t_{\alpha/2}$ mbështetet në $df = n - 1$ shkallë lirie.

Testi i diferencës së çiftëzuar për diferencën e mesatareve

Në qoftë se popullimi i diferecave është normal, **hedhim poshtë H_0: $\mu_d = D_0$ me nivel rëndësie α** (probabiliteti i gabimit të Tipit I) atëherë dhe vetëm atëherë kur vlen **kushti pikësor** ose, në mënyrë ekuivalente, në qoftë se p-vlera është më e vogël se α.

Alternativa	**Hedh poshtë H_0 nëse:**	**p-vlera**
$H_a: \mu_d > D_0$	$t > t_\alpha$	Syprina nën t - shpërndarjen në të djathtë të t
$H_a: \mu_d < D_0$	$t < -t_\alpha$	Syprina nën t - shpërndarjen në të majtë të t
$H_a: \mu_d \neq D_0$	$\lvert t \rvert > t_{\alpha/2}$, dmth. $t > t_{\alpha/2}$ ose $t < -t_{\alpha/2}$	Dyfishi i syprinës nën t - shpërndarjen në të djathtë të $\lvert t \rvert$

Statistika e testit

$$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

t_α, $t_{\alpha/2}$ dhe p-vlerat mbështeten në $n - 1$ shkallë lirie.

Shembull: Intervali dhe testi i diferencave të çiftëzuara

Vetura	Garazha 1	Garazha 2	Diferenca
Vet. 1	$ 7.10	$ 7.90	-0.8
Vet. 2	9.00	10.10	-1.1
Vet. 3	11.00	12.20	-1.2
Vet. 4	8.90	8.80	0.1
Vet. 5	9.90	10.40	-0.5
Vet. 6	9.10	9.80	-0.7
Vet. 7	10.30	11.70	-1.4
		Mesatarja	-0.8
		Dev Std.	0.5033

Tabela 9.3

95% Intervali i besueshmërisë

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = -0.8 \pm 2.447 \left(\frac{0.5033}{\sqrt{7}} \right) = -0.8 \pm 0.4654 = [-1.2654, -0.3346]$$

9.5 Intervali i mostrave të mëdha për diferencën e proporcioneve

Në qoftë se dy mostra të pavarura janë të dyja të mëdha, **100(1 - α)% intervali i besueshmërisë për p_1 - p_2** është

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 - 1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2 - 1}}$$

Testet e mostrave të mëdha për diferencën e proporcioneve

Në qoftë se dy populime të zgjedhura janë të dyja të mëdha, **hedhim poshtë H_0: $p_1 - p_2 = D_0$ me nivel rëndësie α** atëherë dhe vetëm atëherë kur vlen vetia përkatëse pikësore, ose, në trajtë ekuivalente, kur p-vlera përkatëse është më e vogël se α.

Alternativa	**Hedh poshtë H_0 nëse:**	**p-vlera**
$H_a: p_1 - p_2 > D_0$	$z > z_\alpha$	Syprina nën z-shpërndarjen në të djathtë të z
$H_a: p_1 - p_2 < D_0$	$z < -z_\alpha$	Syprina nën z-shpërndarjen në të majtë të z
$H_a: p_1 - p_2 \neq D_0$	$\lvert z\rvert > z_{\alpha/2}$, dmth. $z > z_{\alpha/2}$ ose $z < -z_{\alpha/2}$	Dyfishi i syprinës nën z-shpërndarjen në të djathtë të $\lvert z\rvert$

Statistika e testit

$D_0 = 0$:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$D_0 \neq 0$:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 - 1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2 - 1}}}$$

Shembull: Intervali dhe testi i diferencave të çiftëzuara

Shembull 9.11
Reklamë mediash

$$\hat{p}_1 = \frac{631}{1000}, \quad \hat{p}_2 = \frac{798}{1000}, \quad \hat{p} = \frac{631+798}{1000+1000}$$

95% intervali i besueshmërisë për p_1 - p_2

$$(0.631-0.798) \pm 1.96\sqrt{\frac{0.631(1-0.369)}{1000-1}+\frac{0.798(1-0.202)}{1000-1}} =$$

$$-0.167 \pm 0.0389 = [-0.2059, -0.1281]$$

Testi i H_0: p_1 - p_2 = 0 kundrejt H_a: p_1 - $p_2 \neq 0$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{(0.631-0.798)-0}{\sqrt{0.7145(1-0.7145)\left(\frac{1}{1000}+\frac{1}{1000}\right)}}$$

$$= -8.2673 < 3.29 = -z_{0.0005}, \ p-value = 2 \times P(z < -8.2673) < 0.001$$

Përfundimet statistikore mbështetur në dy mostra

9.1 Krahasimi i dy mesatareve të popullimit duke shfrytëzuar mostra të mëdha të pavarura

9.2 Krahasimi i dy mesatareve të popullimit duke shfrytëzuar mostra të vogla të pavarura

9.3 Krahasimi i dy variancave të popullimit duke shfrytëzuar mostra të pavarura

9.4 Eksperimentet e diferencës së çiftëzuar

9.5 Krahasimi i dy proporcioneve të popullimit duke shfrytëzuar mostra të mëdha të pavarura