

Anvendt Statistik

Lektion 5

Sammenligning af to grupper

- * Sammenligning af middelværdier
- * Sammenligning af andele

Motiverende eksempel

- Antal minutter brugt på rengøring/madlavning:

	Rengøring/Madlavning		
Køn	Stikprøvesørrelse	Gennemsnit	Standardafvigelse
Mænd	1219	23	32
Kvinder	733	37	16

- Der er **to variable** registeret for hver person
 - Køn: **Forklarende** variabel
Binær/dikotom (to mulige værdier)
 - Tid: **Afhængig** variabel
Kontinuert/skala
- Uafhængige grupper (mænd/kvinder)

Standardfejlen for forskellen

- **Mål:**
 - Sammenligne middelværdier μ_1 og μ_2 for to grupper.
- **Ny parameter**
 - Differencen $\mu_2 - \mu_1$ er en parameter
- **Estimat**
 - $\bar{y}_2 - \bar{y}_1$ er et estimat for $\mu_2 - \mu_1$
- **Husk:** Et 95% **konfidensinterval** er af typen
Punktestimat $\pm 2 \cdot se$
hvis punktestimatet er (approks.) **normalfordelt**.

Standardfejlen for $\bar{y}_1 - \bar{y}_2$

- Antag vi har to uafhængige stikprøver, og at se_1 og se_2 er standardfejlen for hhv. y_1 og y_2 . Da er den estimerede standardfejl for $y_1 - y_2$

$$se = \sqrt{se_1^2 + se_2^2}$$

hvor $se_i = s_i / \sqrt{n_i}$.

- **Eksempel:** Oprydning/Madlavning

$$se = \sqrt{se_1^2 + se_2^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(32)^2}{1219} + \frac{(16)^2}{733}} = 1.09$$

Konfidensinterval for $\mu_1 - \mu_2$

- For uafhængige stikprøver fra to grupper der har normale populationsfordelinger er et $(1-\alpha)100\%$ konfidensintervallet for $\mu_1 - \mu_2$ givet ved

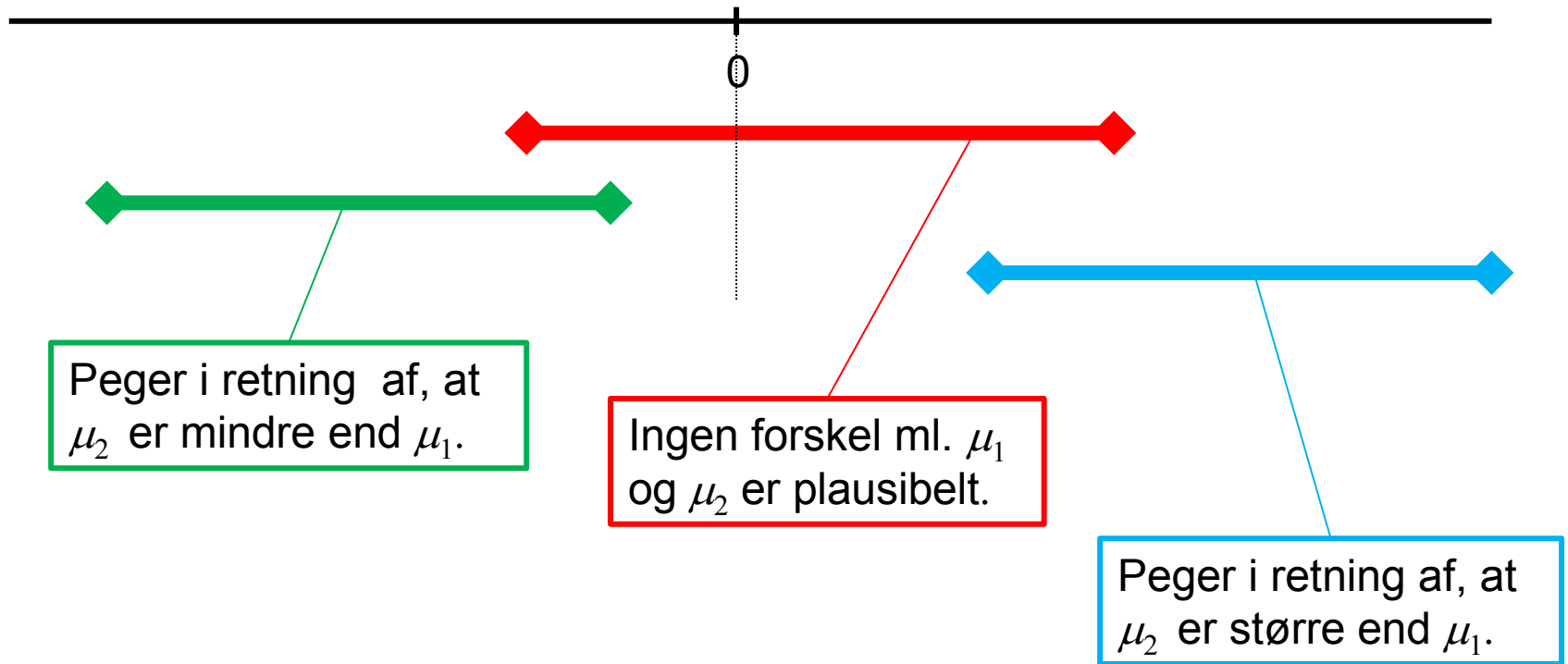
$$(\bar{y}_2 - \bar{y}_1) \pm t_{\alpha/2} \cdot se$$

hvor t har $df = n_1 + n_2 - 2$ frihedsgrader.

- **Eksempel:** Oprydning/Madlavning

Konfidensintervaller

■ Konfidensintervaller for $\mu_2 - \mu_1$



Hypotesetest for $\mu_2 - \mu_1$

- **Antagelser:** Normale populationer

- **Nul-hypotese:**

- $H_0: \mu_2 - \mu_1 = 0$ (ingen forskel)

- **Alternativ hypotese:**

- $H_a: \mu_2 - \mu_1 \neq 0$ (en forskel)

$$t = \frac{\text{punktestimat} - H_0\text{værdi}}{\text{standardfejl}}$$

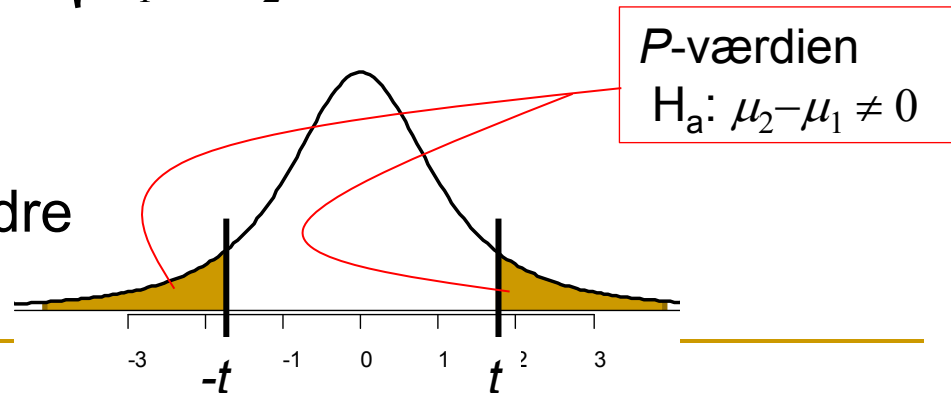
- **Teststørrelse:**

- $t = \frac{(\bar{y}_2 - \bar{y}_1) - 0}{se}$ hvor $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

- **P-værdi**

- **Konklusion:**

- Jo lavere P-værdi jo mindre tror vi på H_0 .



Éksempel

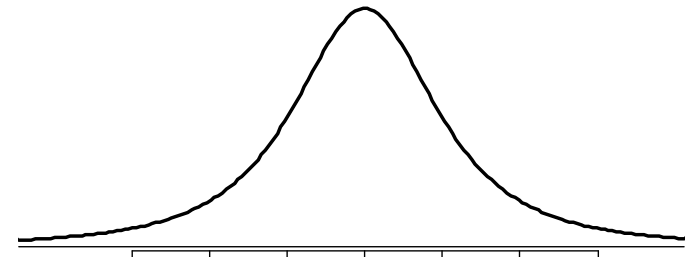
- Antal minutter brugt på rengøring/madlavning:

	Rengøring/Madlavning		
Køn	Stikprøvesørrelse	Gennemsnit	Standardafvigelse
Mænd	$n_1=1219$	$\bar{y}_1=23$	$s_1=32$
Kvinder	$n_2=733$	$\bar{y}_2=37$	$s_2=16$

- Hypoteser: $H_0: \mu_2 - \mu_1 = 0$ vs $H_a: \mu_2 - \mu_1 \neq 0$

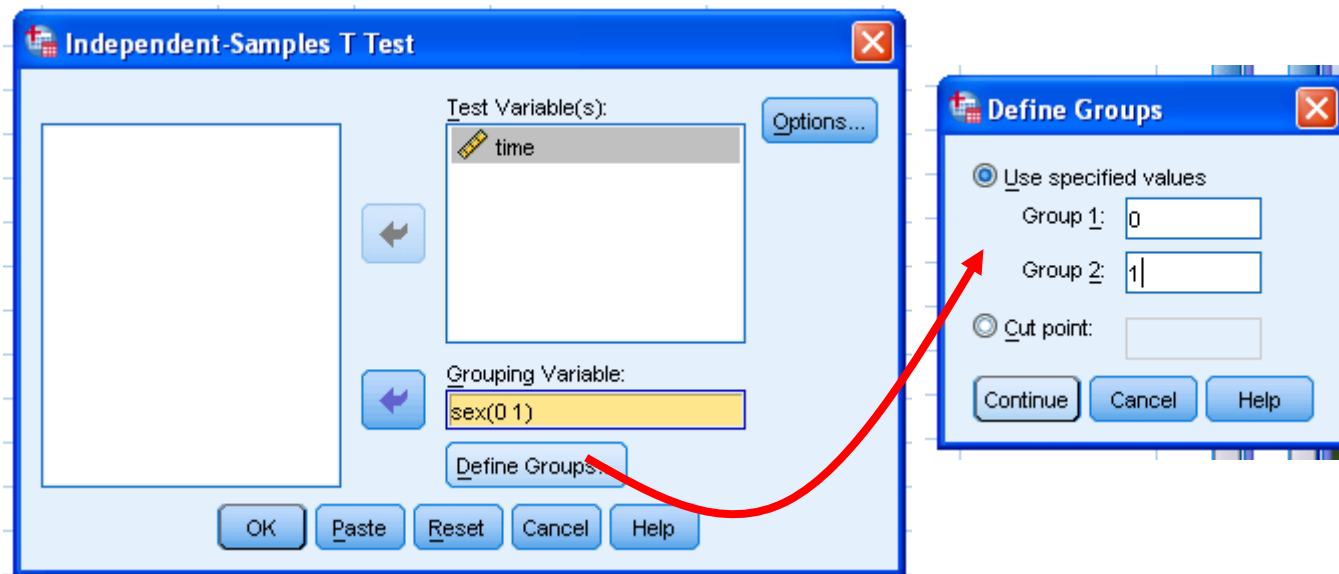
- $$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} =$$

- $$t = \frac{(\bar{y}_2 - \bar{y}_1) - 0}{se} =$$



SPSS

- **SPSS:** Analyze → Compare Means → Independent-Sample T Test
- **Test Variable(s):** Afhængig variabel
- **Grouping variabel:** Forklarende variabel
- **Define Groups:** Angiv hvilke værdier af den forklarende variabel, der svarer til de to grupper.



SPSS Output

■ Opsummering af de to grupper

Group Statistics

sex	N	Mean	Std. Deviation	Std. Error Mean
time Male	1219	22,98	32,002	,917
Female	733	37,01	16,014	,591

■ Test af forskel i middelværdi:

Independent Samples Test

		Levene's Test for Equality of Variances		t		
		F	Sig.	t	df	Sig. (2-tailed)
time	Equal variances assumed	79,892	,000	-11,065	1950	,000
	Equal variances not assumed			-12,861	1896,441	,000

t-teststørrelse

P-værdi
to-sidet test

■ Konfidensinterval:

l)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
			Lower	Upper
0	-14,030	1,268	-16,516	-11,543
0	-14,030	1,091	-16,169	-11,890

Sammenligne μ_1 og μ_2 for afhængige stikprøver

- Typisk eksempel på afhængige grupper, er hvor observationer i de to grupper er **parrede**.
- **Eksempel:** Hver af 28 studerende får målt reaktionstider under bilkørsel under to omstændigheder:
 - 1) Mens de snakker i mobil (gruppe1) $y_{i,1}$
 - 2) Uden de snakker i mobil (gruppe2) $y_{i,2}$
- For hver studerende har vi en forskel i reaktionstid:

$$y_{2,i} - y_{1,i}$$

Sammenligne μ_1 og μ_2 for afhængige stikprøver

- Lad y_1 og y_2 være gennemsnittet for hhv. gruppe 1 og gruppe 2.
- Lad y_d være gennemsnittet af differencerne.
- Der gælder $\bar{y}_d = \bar{y}_2 - \bar{y}_1$
- Dvs. hvis vi vil teste forskelle er det nok at se på gennemsnittet af differencerne.

- Et $(1-\alpha)100\%$ konfidensinterval for $\mu_2 - \mu_1$

$$\bar{y} \pm t_{\alpha/2} \cdot \frac{s_d}{n} \quad df = n-1$$

hvor s_d er standardafvigelsen for differencerne.

Signifikanstest for $\hat{\pi}_2 - \hat{\pi}_1$ (parrede obs.)

- **Antagelser:** Normale populationer

- **Nul-hypotese:**

 - $H_0: \mu_D = 0$ (ingen effekt)

- **Alternativ hypotese:**

 - $H_a: \mu_D \neq 0$ (en effekt)

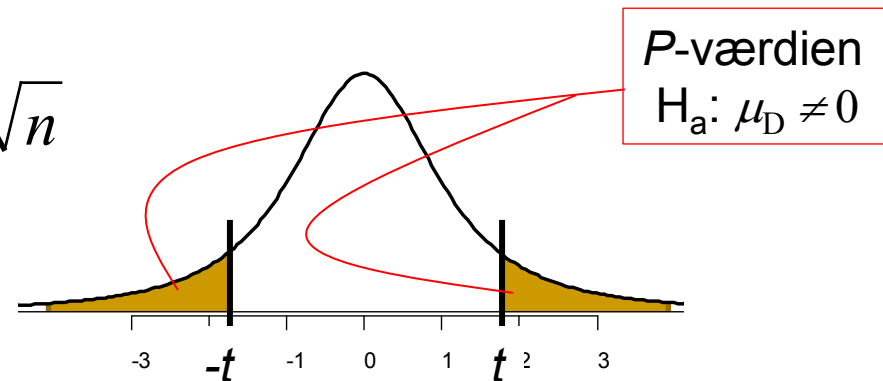
- **Teststørrelse:**

 - $t = \frac{\bar{y}_D - 0}{se}$ hvor $se = s_D / \sqrt{n}$

- **P-værdi**

- **Konklusion:**

 - Jo lavere P -værdi jo mindre tror vi på H_0



Eksempel

- Hypoteser

$$H_0: \mu_D = 0 \quad \text{vs} \quad H_a: \mu_D \neq 0$$

- Gennemsnit

$$\bar{y}_D = -50.625$$

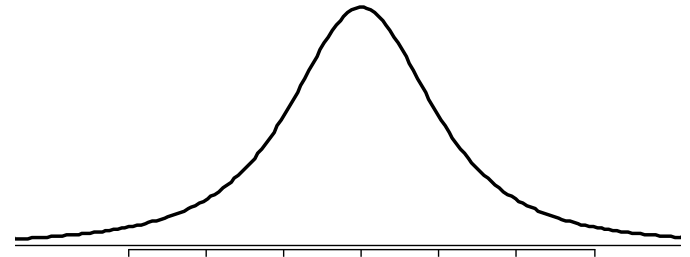
- Standardafvigelse

$$s_D = 50.486$$

- Teststørrelse

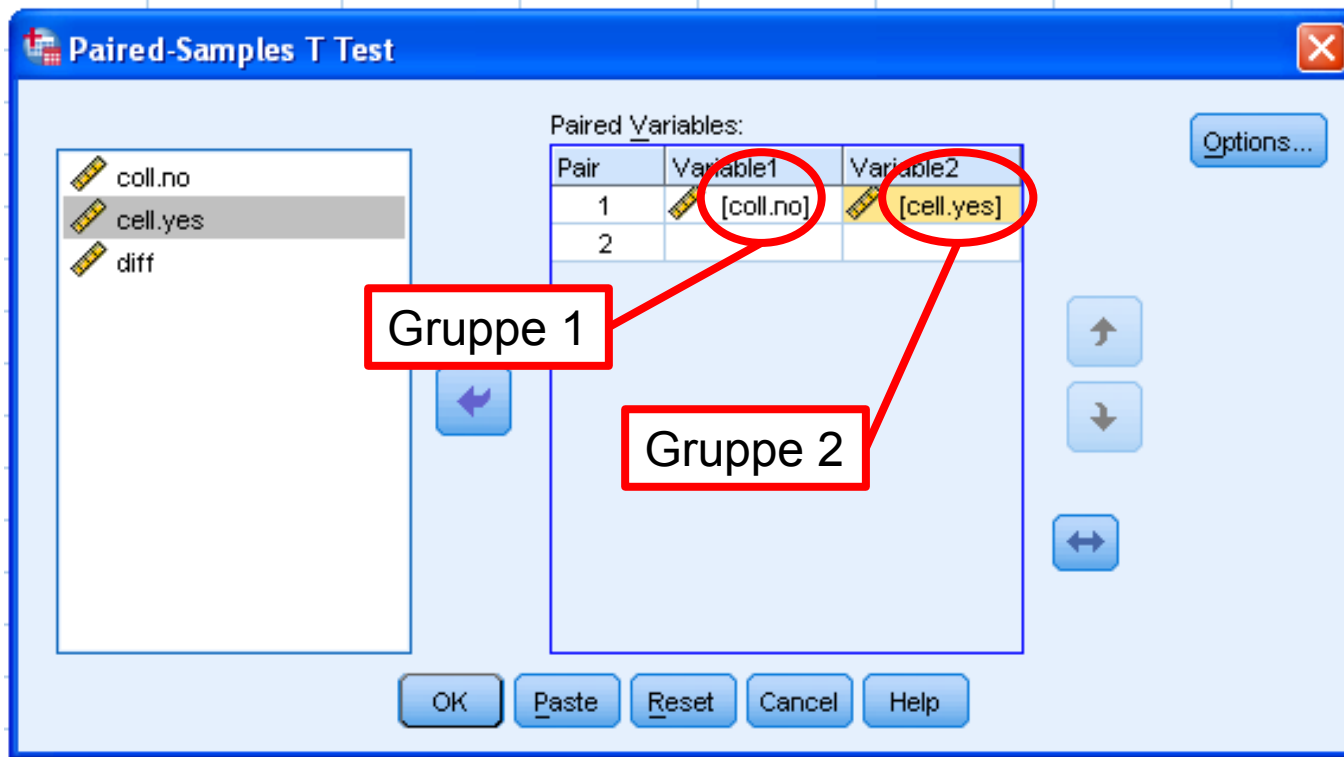
$$t = \frac{\bar{y}_D - 0}{s_D / \sqrt{n}} =$$

Uden mobil	Med mobil	Diff.
604	636	32
556	623	67
540	615	75
522	672	150
459	601	142
544	600	56
:	:	:



SPSS

- **SPSS:** Analyze → Compare Means → Paired-Samples T Test



SPSS: Resultat

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 coll.no - cell.yes	-50,625	52,486	9,278	-69,548	-31,702	-5,456	31	,000

$$se = s_d / \sqrt{n} = 52.486 / \sqrt{32} = 9.28$$

$$t = \frac{\bar{y}_d - 0}{se} = \frac{-50.825}{9.28} = -5.455$$

Test direkte på differencerne

- Lav et t -test af differencerne

One-Sample Test						
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
diff	5,456	31	,000	50,625	31,70	69,55

- Bemærk at t -er præcis som før og dermed er P -værdien som før.

Test af forskel i andele

- Effekten af bøn på udfald af operation:

Bøn	Komplikationer	Ej komplikationer	Total	Andel kompl.
Ja	315	289	604	0.522
Nej	304	293	597	0.509

- Der er **to variable** registeret for hver person
 - Bøn: **Forklarende** variabel
Binær/dikotom (to mulige værdier)
 - Udfald: **Afhængig** variabel
Binær/dikotom (to mulige værdier)
- Uafhængige grupper (Bøn/Ej bøn)

Standardfejlen for $\hat{\pi}_2 - \hat{\pi}_1$

- Standardfejlen for $\hat{\pi}_2 - \hat{\pi}_1$ er

$$se = \sqrt{se_1^2 + se_2^2}$$

hvor $se_i = \sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)/n_i}$.

- **Eksempel:** Bøn og operation

$$\begin{aligned} se &= \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}} \\ &= \sqrt{\frac{0.522(1 - 0.522)}{604} + \frac{0.509(1 - 0.509)}{597}} = 0.0288 \end{aligned}$$

Konfidensinterval for $\pi_2 - \pi_1$

- For store stikprøver er et $(1-\alpha)100\%$ konfidensinterval for forskellen $\pi_2 - \pi_1$ mellem to populationer

$$(\hat{\pi}_2 - \hat{\pi}_1) \pm z_{\alpha/2} \cdot se$$

- **Eksempel:** Bøn og operation

- Et 95% konfidensinterval for forskellen i andele:

$$\begin{aligned}(\hat{\pi}_2 - \hat{\pi}_1) \pm z_{\alpha/2} \cdot se &= (0.509 - 0.522) \pm 1.96 \cdot 0.0288 \\ &= -0.013 \pm 0.057 \quad (-0.07 ; 0.04)\end{aligned}$$

- Da KI'et indeholder 0, er "ingen forskel" plausibelt.

Signifikanstest for $\hat{\pi}_2 - \hat{\pi}_1$

- **Antagelser:** Store stikprøver
- **Nul-hypotese:**
 - $H_0: \pi_2 - \pi_1 = 0$ (ingen effekt)

$$z = \frac{\text{punktestimat} - H_0 \text{ værdi}}{\text{standardfejl}}$$

- **Alternativ hypotese:**
 - $H_a: \pi_2 - \pi_1 \neq 0$ (en effekt)

- **Teststørrelse:**

$$\square z = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - 0}{se_0} \quad \text{hvor} \quad se_0 = \sqrt{\hat{\pi}(1 - \hat{\pi}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

- **P-værdi**

- **Konklusion:**

- Jo lavere P -værdi jo mindre tror vi på H_0

$\hat{\pi}$ er den overordnede andel, når grupper ignoreres.

Test af forskel i andele

- Effekten af bøn på udfald af operation:

Bøn	Komplikationer	Ej komplikationer	Total	Andel kompl.
Ja	315	289	$n_1=604$	$\hat{\pi}_1=0.522$
Nej	304	293	$n_2=597$	$\hat{\pi}_2=0.509$

- **Hypoteser:** $H_0: \pi_2 - \pi_1 = 0$ vs $H_a: \pi_2 - \pi_1 \neq 0$

- $$\hat{\pi} = \frac{315 + 304}{604 + 597} = 0.515$$

- $$se_0 = \sqrt{\hat{\pi}(1 - \hat{\pi}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} =$$

- $$Z = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - 0}{se_0} =$$

