121129 Försöksplanering (ex) Plastgjutning

M	A	MA
+	−	−
+	+	+
+	−	−
+	+	+
+	−	−
+	+	+
+	−	−
+	+	+

A	A	AA
−	−	+
+	+	+
−	−	+
+	+	+
−	−	+
+	+	+
−	−	+
+	+	+

Användbara räkneregler

A	B	C	AB	AC	BC	D ABC

Generator $D = ABC$

$I = D \cdot D = ABCD$ DEFINIERAD RELATION!

$A \cdot I = \cancel{A} \cdot \cancel{A}BCD = BCD$

$B \cdot I = \cancel{B} \cdot A\cancel{B}CD = ACD$

$AC \cdot I = \cancel{AC} \cdot \cancel{A}B\cancel{C}D = BD$

Antal bokstäver i det kortaste ordet kallas upplösningen. Här står det fyra, skrivs IV.

2^{5-2} försök. Varje faktor har två nivåer, varje nivå har fem faktorer. Men här har man reducerat två gånger.

Generatorer $D=ABC$ och $E=BC$, definierande relationer.

$I_1 = ABCD$ $\quad I_3 = I_1 \cdot I_2 = ABCD \cdot BCE = ADE$

$I_2 = BCE$ $\quad$ Vad är upplösningen? III

$A \cdot I_1 = A \cdot ABCD = BCD$

$A \cdot I_2 = ABCE$

$A \cdot I_3 = A \cdot ADE = DE$

$l_A \longrightarrow A + DE + BCD + ABCE$

Upplösning III medför att huvudeffekter och 2-faktorsamspel sammanblandas.

Upplösning IV

$I = ABCD$

$A \cdot I = A \cdot ABCD = BCD$

$AB \cdot I = AB \cdot ABCD = CD$

Upplösning IV medför att 2-faktorsamspel sammanblandas med varandra

Upplösning V

$I = ABCDE$

$A \cdot I = A \cdot ABCDE = BCDE$ osv...

$AB \cdot I = AB \cdot ABCDE = CDE$ osv...

2^{6-3} försök A B C AB $\overset{D}{AC}$ $\overset{E}{BC}$ $\overset{F}{ABC}$

Generatorer $D = AC$ $E = BC$ $F = ABC$

Definierade relationer.

$I_1 = ACD$

$I_2 = BCE$

$I_3 = ABCF$

$I_4 = I_1 \cdot I_2 = ACD \cdot BCE = ABDE$

$I_5 = I_1 \cdot I_3 = \cancel{A}\cancel{C}D \cdot \cancel{A}B\cancel{C}F = BDF$

$I_6 = I_2 \cdot I_3 = \cancel{B}\cancel{C}E \cdot A\cancel{B}\cancel{C}F = AEF$

$I_7 = I_1 \cdot I_2 \cdot I_3 = \cancel{A}\cancel{C}D \cdot \cancel{B}\cancel{C}E \cdot \cancel{A}\cancel{B}CF = CDEF$

Upplösning III

Har man tvåfaktorsamspel som är viktigt då bör man inte reducera så mycket

Ex Kemisk process, 2^3 försök

T = temperatur, P = tryck och K = katalysator 3 faktorer

T	P	K	TP	TK	PK	TPK	y
−	−	−	+	+	+	−	55
+	−	−	−	−	+	+	52
−	+	−	+	+	−	+	59
+	+	−	−	−	−	−	61
−	−	+	−	−	−	+	51
+	−	+	+	+	+	−	48
−	+	+	−	−	+	−	65
+	+	+	+	+		+	66

$l_M = 57{,}125$ $l_T = -0{,}75$ $l_P = 11{,}25$ $l_K = 0{,}75$ $l_{TP} = 2{,}25$

$l_{TK} = -0{,}25$ $l_{PK} = 4{,}75$ $l_{TPK} = -0{,}25$

Dessa försök behålls

2^{3-1} plan

T	P	K	TP	TK	PK	TPK	y
+	−	−	−	−	+	+	52
−	+	−	−	+	−	+	59
−	−	+	+	−	−	+	51
+	+	+	+	+	+	+	66

$l_T = \frac{1}{2}(52+66) - \frac{1}{2}(59+51) = 4,0 = \underbrace{-0,75 + 4,75}_{\text{från det fullständiga försöket}} = 4,0$

$l_P = \frac{1}{2}(59+66) - \frac{1}{2}(52+51) = 11,0 = 11,25 - 0,25 = 11,0$

$l_{TPK} = \frac{1}{4}(52+59+51+66) = 57,0$

$57,125 - \frac{0,25}{2} = 57,0$

Medelvärdet divideras med 8 medan TPK divideras med 4, och det måste regleras genom att de båda divideras med samma tal.

121203 Test och variansanalys

Styrka = $1-\beta$

$P(\text{förkasta } H_0 \mid H_0 \text{ falsk})$

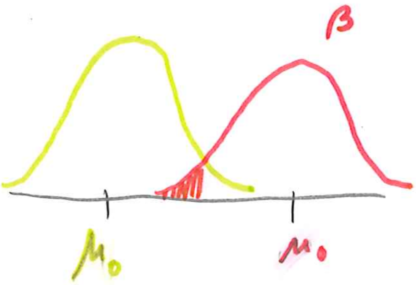
	H_0 sann	H_1 sann
Vi tror H_0	OK	Typ II
Vi tror H_1	Typ I	OK

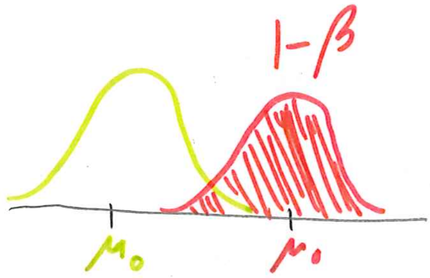
$P(\text{Vi tror på } H_0 \mid H_0 \text{ sann}) = 1-\alpha$

$P(\text{Vi tror på } H_1 \mid H_0 \text{ sann}) = \alpha$ (Typ I-fel)

$P(\text{Vi tror på } H_0 \mid H_1 \text{ sann}) = \beta$ (typ II-fel)

$P(\text{Vi tror på } H_1 \mid H_1 \text{ sann}) = 1-\beta$ (styrkan)





Om α är liten, så är β stor och tvärtom. Man måste optimera för det som ger tillräckligt låg signifikansnivå α och så stor styrka $(1-\beta)$ som möjligt.

Exempel "styrka"

Vi har $\xi = N(\mu, 20)$. Tar ett urval om $n=100$ och undersöker om $\mu > 60$. Stickprovet gav $\bar{x} = 63{,}40$

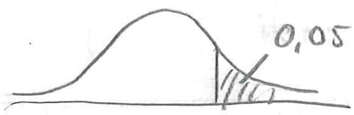
Testar på signifikansnivån $\alpha = 0{,}05$ om $\mu > 60$

Testvariabel $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{63{,}40 - 60}{20/\sqrt{100}} = 1{,}7$

Stickprovet

$H_0 = \mu \leq 60$

$H_1 = \mu > 60$



Vårt $\alpha = 0{,}05$ ger ett kritiskt värde $x^c = 1{,}645$ från tabell. Vår testvariabel är större än det kritiska värdet: $z > x^c$. Vi förkastar H_0. Vår undersökning tyder på att $\mu > 60$.

Om nu $\mu = 63$. Vad är då sannolikheten att göra fel, dvs $P(\text{vi tror på } H_0 \mid H_1 \text{ sann}) = \beta$

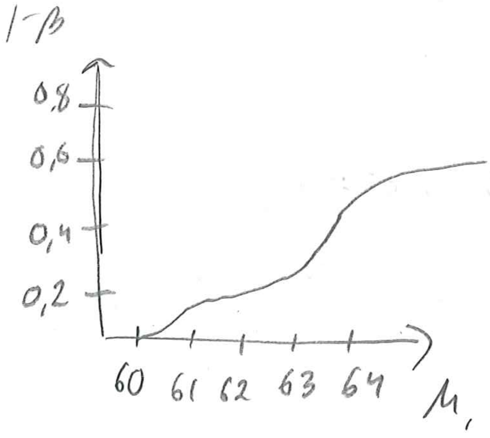
Vi beräknar $x^c = \mu_0 + z^c \cdot \dfrac{s}{\sqrt{n}} = 60 + 1{,}645 \cdot \dfrac{20}{\sqrt{100}} = 63.29$

Om nu $\bar{x}$ blir större än eller lika med x^c förkastar vi H_0 och tror H_1 är sann.

Vi beräknar först β och tar sedan $1-\beta$ för att räkna ut testets styrka.

$\beta = P(\bar{x} \leq 63{,}29 \mid \mu = \mu_1) = P\left(z \leq \dfrac{63{,}29 - 63}{20/\sqrt{100}}\right) = 0{,}5577 \implies$

Styrkan $= 1 - \beta = 1 - 0{,}5577 = 0{,}4423$



μ_1 väljs och sedan plottas μ_1 mot styrkan

Sambandet mellan α, β och n

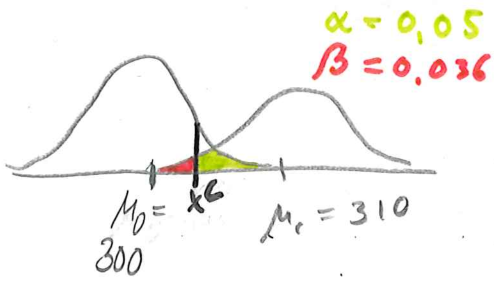
Det finns ett samband mellan α, β och n

Ex Antag $X \sim N(\mu, 30)$

Testa H_0, $\mu = 300$

H_1, $\mu = 310$

Bestäm n så att $\alpha = 0{,}05$, $\beta = 0{,}036$



Vi behöver veta vilket x^c som ger ovanstående avgränsning samt vilket n som behövs

μ_0 och α: $x^c = \mu_0 + z_0^c \cdot \frac{\sigma}{\sqrt{n}} = 300 + 1{,}645 \cdot \frac{30}{\sqrt{n}}$

μ_1 och β: $x^c = \mu_1 + z_1^c \cdot \frac{\sigma}{\sqrt{n}} = 310 - 1{,}8 \cdot \frac{30}{\sqrt{n}}$

Vi sätter uttrycken lika och löser ut n:

$$300 + 1{,}645 \cdot \frac{30}{\sqrt{n}} = 310 - 1{,}8 \cdot \frac{30}{\sqrt{n}} \Rightarrow n = 106{,}8 \approx 107$$

Kritiskt värde $x^c = \mu_0 + z_0^c \cdot \frac{\sigma}{\sqrt{n}} = 300 + 1{,}645 \cdot \frac{30}{\sqrt{106{,}8}} = 304{,}78$

Räkneexempel 1.10

Bildäck, livslängden beskrivs som $N(\mu = 1510, \sigma = 35)$

Ny process skall göra däcken mer hållbara (dvs längre livslängd).

a) Antag att nya processen ger en livslängd på 1525. Hur många däck måste tillverkas om man vill ha en signifikansnivå $\alpha = 0{,}01$ och β-risk $= 0{,}03$

Vi har redan bestämt α och β och behöver räkna ut x^c mha dessa. Vi utgår från följande uttryck för x^c under antagande om våra α och β:

① $x^c = \mu_0 + z_0^c \cdot \frac{\sigma}{\sqrt{n}} = 1510 + 2{,}33 \cdot \frac{35}{\sqrt{n}}$

② $x^c = \mu_1 + z_1^c \cdot \frac{\sigma}{\sqrt{n}} = 1525 - 1{,}88 \cdot \frac{35}{\sqrt{n}}$

Vi sätter ① = ②

$$1510 - 2{,}33 \cdot \frac{35}{\sqrt{n}} = 1525 - 1{,}88 \cdot \frac{35}{\sqrt{n}} \Rightarrow n = 96{,}5 \approx 97$$

b) Beräkna testets styrka. $1-\beta$ om $\mu = 1510$ och 50 däck tillverkas.

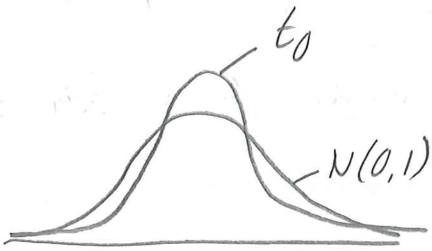
$$\text{Styrkan} = (1-\beta) = P(\bar{x} > \underbrace{1510 + 2{,}33 \cdot \frac{35}{\sqrt{50}}}_{1521}) = 1 - P\left(z \leq \frac{1521 - 1520}{35/\sqrt{50}}\right) =$$

$$= 1 - P(z \leq 0{,}31) = 1 - 0{,}6217 = 0{,}3783$$

t-fördelningen

$\xi \sim N(\mu, \sigma)$ σ är okänd

$$\frac{\xi - \mu}{\sigma/\sqrt{n}} \sim t_{n-1}$$



Hittils har test varit på $\xi \sim N(\mu, \sigma)$ med σ = känd
Om σ är okänt kan vi inte använda normalfördelningen direkt. Drar vi ett litet urval måste vi använda t-fördelningen.

När man genomför ett t-test gör man på samma sätt som för normalfördelningstest med skillnaden att man använder tabellen för t-fördeln. för att få sitt kritiska värde.

Ensidiga test:

H_0 $\mu > \mu_0$

H_1 $\mu < \mu_0$

Dubbelsidiga test

H_1 $\mu \neq \mu_0$

Test av populationsproportion, p

Om vi har ett stort urval kan fördelningen av populationsproportionen approximeras som normalfördelad med väntevärdet p, och standardavvikelsen $\sqrt{\frac{p(1-p)}{n}}$

Som tumregel kan man använda $p(1-p)n > 10$.

Notera att populationsprop. ALDRIG approximeras med t-fördelningen.

Vid test räknar man som om H_0 vore sann dvs både väntevärde och standardavvikelsen är känd.

Testvariabel $$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

Ex 1.10 Ett slumpmässigt urval om 200 pers har köpt en viss produkt och 87 var kvinnor.

Är det egentligen lika fördelad mellan könen?

Vi testar på signifikansnivå $\alpha = 0{,}01$ och beräkna p-värdet.

1. $H_0: p = 0{,}5$
   $H_1: p \neq 0{,}5$ dubbelsidigt

2. $\alpha = 0{,}01$

3. $$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

4. Från urvalet fick vi
   $$p_{kvinnor} = \frac{87}{200} = 0{,}435$$
   $$z = \frac{0{,}435-0{,}5}{\sqrt{\frac{0{,}5(1-0{,}5)}{200}}} = -1{,}84$$

Vi räknar ut p-värdet:

p-värdet $= 2 \cdot P(Z < -1{,}84) =$

$2 \cdot (1 - P(Z < 1{,}84)) = 2 \cdot (1 - 0{,}967) = 0{,}0658$

p-värdet $> \alpha$ så vi kan ej förkasta H_0

Dvs att undersökning antyder att vi inte kan se (på 1% signifikans) att produkten mer av något av könen.

121206 Repitition Hypotestest

Ex. 1.3 En kaffebar säljer i genomsnitt 320 koppar kaffe om dagen med en standardavvikelse på 40 koppar. Efter en annonskampanj har man under de senaste 7 dagarna sålt 2450 koppar. De vill se om annonskampanj haft effekt. Antag genomsnittligt antal koppar är normalfördelad.

a) Ägaren accepterar en signifikansnivå på 5%. Bör man förkasta hypotesen att försäljningen i genomsnitt är oförändrad?

Vi vill testa:

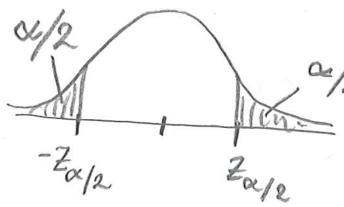
$H_0 : \mu = 320$

$H_1 : \mu \neq 320$

Signifikansnivå $\alpha = 0{,}05$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \Rightarrow Z = \frac{350 - 320}{40/\sqrt{7}} = 1{,}98$$

Jämför Z med $Z_{\alpha/2}$



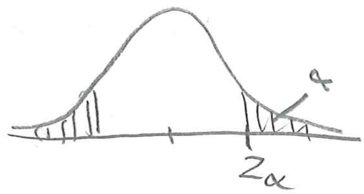
$Z_{\alpha/2} = [\text{fr. tabell}] = 1{,}96$

Vårt $Z = 1{,}98 > Z_{\alpha/2} = 1{,}96 \Rightarrow$ Vi förkastar H_0, medelvärdet har förändrats

b) Vi vill se om det skett en ökning.

Steg 1. $H_0 : \mu \leq 320$

$H_1 : \mu > 320$



Steg 2. Signifikansnivå 5%. $\alpha = 0{,}05$

$\Rightarrow z_\alpha = 1{,}645$

Steg 3. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$

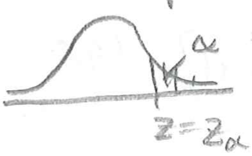
Steg 4. $z = \dfrac{350 - 320}{40/\sqrt{7}} = 1{,}98$

Steg 5. Jämför $z = 1{,}98$ med $z_\alpha = 1{,}645$

$z = 1{,}98 > z_\alpha = 1{,}645 \Rightarrow$ Vi förkastar H_0

c) Antag att man observerar försäljningen under 25 dagar. Vilken genomsnittsförsäljning hade då behövts per dag för att kunna påvisa en ökad försäljning med 1% signifikansnivå.

Vi söker $\bar{x}$ så att p-värdet för testvariabeln blir lika med $\alpha = 0{,}01$



, $P(\xi > \bar{x}) = 0{,}01$

Lättare att räkna ut komplementet $P(\xi \leq \bar{x}) = 0{,}99$

Standardnormalfördela:

$$P\Big(\underbrace{\frac{\xi - \mu}{\sigma/\sqrt{n}}}_{z} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\Big) = P\Big(z \leq \frac{\bar{x} - 320}{40/\sqrt{25}}\Big) = 0{,}99$$

Från tabell ser vi att $P(z \leq 2{,}33) = 0{,}99 \Rightarrow$

$$\frac{\bar{x} - 320}{40/\sqrt{25}} = 2{,}33 \Rightarrow \bar{x} = 338{,}64$$

Det krävs ett genomsnitt på 338,64 koppar per dag för att vi ska kunna se en ökning på 1% signifikansnivå över 25 dagar.

Ex 1.4 Man vill undersöka hållfastheten hos ett material och får mätvärden som är normalfördelat med känt $\sigma = 2,5$. Om väntevärdet överstiger 40 är materialet användbart. Man vågar ta en risk på 0,1 %. Ett stickprov gav $\bar{x} = 39,8$ baserat på $n = 16$ mätningar. Kan man godta materialet?

Steg 1. $H_0 : \mu \geq 40$

$H_1 : \mu < 40$

Steg 2. $\alpha = 0,001$  $\Rightarrow$ kritiskt värde $Z_\alpha = -3,1$

0,001

Steg 3. $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

steg 4. $Z = \dfrac{39,8 - 40}{2,5/\sqrt{16}} = -0,32$

Steg 5. Jämför Z med Z_α

$Z = -0,32 > Z_\alpha = -3,1 \Rightarrow$ Vi kan ej förkasta H_0!

Därför att $Z < Z_\alpha$ för att kunna förkasta H_0, då vi har negativa värden.

P-värdet för testet ges av $p = P(Z \leq -0,32) = 1 - P(Z \leq 0,32) =$
$1 - 0,6255 = 0,3745$

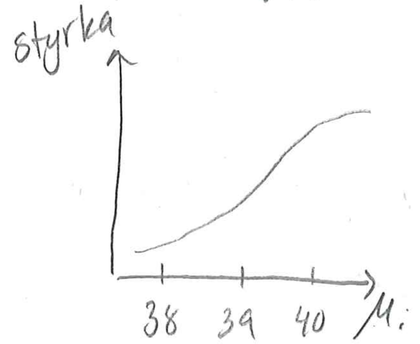
$p = 0,3745$

Vad blir testets styrka om den sanna fördelningen för stickprovet är $\mu_1 = 38$

$\Longrightarrow$

$P(\text{vi förkastar } H_0 \mid H_0 \text{ falsk}) = P(\bar{x} < 38 \mid \mu = \mu_1) = 1 - \beta$

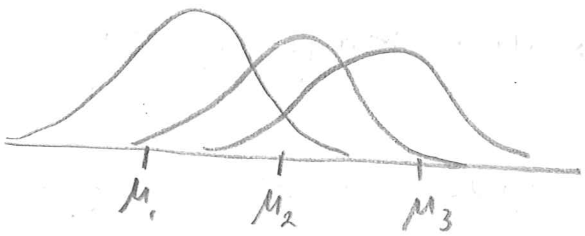
$$P\left(Z \leq \frac{39{,}8 - 38}{2{,}5/\sqrt{16}}\right) = P(Z \leq 2{,}88) = 0{,}9980$$



Variansanalys

Analysis of variance - ANOVA

Jämföra flera "grupper", testa om väntevärderna är olika.



Om endast två grupper jämförs, använd skillnaden i medelvärde i förhållande till osäkerheten.

$$\frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} - \frac{1}{n_2}}}$$

$H_0: \mu_1 = \mu_2$ $\iff$ $H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 - \mu_2 \neq 0$

Handlar det om fler än två kan man inte göra så här med medelvärdesskillnaden.

$\Longrightarrow$

Då vill vi testa:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$

H_1: något $\mu_i \neq \mu_j$, $i \neq j$ (alla måste inte skilja)

Förutsättning för ANOVA är en linjär modell där våra stokastiska variabler $\xi_i \sim N(\mu_i, \sigma)$, $i = 1, \ldots k$

Alla ξ_i är oberoende! Samma standardavvikelse

Varför jämför man bara inte medelvärdena två i taget? Vill vi jämföra 5 olika nivåer får vi 10 olika test. två av testen blir tex:

$\mu_1 = \mu_2$ och $\mu_2 = \mu_3$ Då μ_2 ingår i båda blir testen inte oberoende, vilket gör att det inte går att fastställa en signifikansnivå för den övergripande jämförelsen.

Hade testen varit oberoende hade vi på $\alpha = 0{,}05$ haft totalt $1 - 0{,}95^k$ sannolikhet att felaktigt förkasta någon av nollhypotesena. I fallet $k = 10$: $1 - 0{,}95^{10} = 0{,}401$ sannolikhet att förkasta någon av nollhypoteserna.

Vi behöver variansanalys!

För att kunna göra variansanalys krävs att:

1. Urvalen från de k grupperna är oberoende!
2. Observationerna är normalfördelade, eventuellt med olika medelvärden men samma varians.

Variansanalys har formen av ett (eller flera) hypotest där testvariabeln är en kvot mellan två varianser som följer en F-fördelning.

Metoden ANOVA har fem huvudsteg:

1. Definiera hypoteserna

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$

H_1: något $\mu_i \neq \mu_j$, $i \neq j$ (alla kan vara olika också)

2. Bestäm signifikansnivå

F-test är ensidigt och förkastelseområdet är till höger i övre svansen.

Ofta finns bara tabell för $\alpha = 0{,}05$ och F-fördeln. beror på antalet frihetsgrader för nämnare och täljare i testvariabel.

3. $F = \frac{MSB}{MSE}$, mean square blocks

 " error

Testvariabeln kommer följa en F-fördeln. med två frihetsgradsvärden. Ett för täljaren, ett för nämnaren. Jämför vi k medelvärden har vi (k-1) frihetsgrader för täljaren. Nämnaren frihetsgradstal beror på flera faktorer (egentligen antalet frihetsgrader för måttelet).

4. Tag stickprov för alla grupper. Fyll i en ANOVA-tabell, med hjälp av denna tabell beräknas F-värdet som vi sedan jämför med F-fördelningens kritiska värde för att avgöra om det finns någon effekt/skillnad mellan μ_1 och μ_K.

5. Förkasta H_0 om vårt framräknande F-värde ligger i förkastelseområdet dvs om $F > F^{k}_{\nu_1, \nu_2}$

ANOVA-tabell exempel

Variationskälla	SS Sum of squa	df frihetsgrader	MS medelsumm	F_0
Mellan grupper, behandlingar	SS_B	$K-1$	MS_B	$\frac{MS_B}{MS_E}$
Okänd fel/error inom gruppen	SS_E	$N-k$	MS_E	
Totalt	SS_T	$N-1$		

Vi har data för:
k grupper
n observationer/grupp
Totalt $N = k \cdot n$

$SS_B = B - C$

$SS_E = A - B$

$SS_T = A - C$

$$A = \sum_{j=1}^{k} \cdot \sum_{i=1}^{n_j} \cdot x_{ij}^2$$

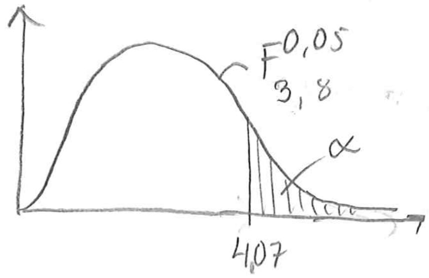
$$B = \frac{\sum_{j=1}^{k} \cdot \left(\sum_{i=1}^{n_j} \cdot x_{ij}\right)^2}{n_j}$$

$$C = \frac{\left(\sum_{j=1}^{k} \cdot \sum_{i=1}^{n_j} \cdot x_{ij}\right)^2}{N}$$

För att kvadratsummorna SS_B och SS_E skall kunna jämföras sätter vi dem i relation till antalet frihetsgrader (df). Detta ger oss, MS, medelkvadratsumma. Förhållandet i variation $\frac{MS_B}{MS_E}$ används som testvariabel, F_0. Om H_0 är som så är $F_0 = \frac{MS_B}{MS_E} \sim F_{r_1=k-1,\, r_2=N-k}$

Förkasta $H_0 : \mu_1 = \ldots = \mu_k$

Ensidigt test: då värdet på testvariabeln är i förkastelseområdet.



$$F\begin{pmatrix}\alpha = 0{,}05 \\ r_1 = 3 \\ r_2 = 8\end{pmatrix} = 4{,}07$$

Alla n_j behöver ej vara lika, men testet har som bäst styrka när alla n_j är lika.

Ensidig (envägs) variansanalys

Analys av en förklarande variabel på flera nivåer. Tex livslängden hos lågenergilampor tillverkade i olika produktionsenheter.

Livslängd som funktion av produktsenhet

$\uparrow$ Förklarande variabel

faktor

$\left.\begin{matrix}1\\2\\3\end{matrix}\right\}$ Nivåer

Här använder man då ANOVA-tabellen

Exempel ANOVA

Livslängd hos lågenergilampor som funktion av produktionsenhet.

Data:

Produktionsenhet	livslängd (h)
1	1000, 1100, 1200
2	1000, 900, 1000
3	1300, 1000, 1100

① $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : alla ej lika

② Vi jämför 3 grupper så frihetsgradtalet för täljaren (MS_B) är $3-1=2$

Totalt finns 9 observationer så nämnaren (MS_E) frihetsgradtal är $9-3=6$.

Vårt kritiska F-värde blir från tabellen $F^{\alpha=0{,}05}_{v_1=2,\, v_2=6} = 5{,}14$

③ Välj testvariabel

$$F = \frac{MS_{produktionsenhet}}{MS_{okänd}} =$$

④ Vårt urval ger $A = \sum_{j=1}^{3} \sum_{i=1}^{3} \cdot x_{ij}^2 = 1000^2 + 1100^2 + 1200^2 +$

$1000^2 + 900^2 + 1000^2 +$

$1300^2 + 1000^2 + 1100^2 =$

$= 10\ 360\ 000$

$$B = \sum_{j=1}^{3} \cdot \frac{\left(\sum_{i=1}^{3} \cdot x\right)^2}{n_j} = \frac{1}{3} \cdot (1000+1100+1200)^2 + \frac{1}{3}(1000+900+1000)^2 +$$

$$\frac{1}{3} \cdot (1300+1000+1100)^2 = 10\ 286\ 666{,}67$$

$$C = \frac{\left(\sum_{i=1}^{3} \sum_{j=1}^{3} X_{ij}\right)^2}{N} = \frac{1}{9}(1000+1100+1200+1000+900+1000+1300+1000+1100)^2 = 10\ 240\ 000$$

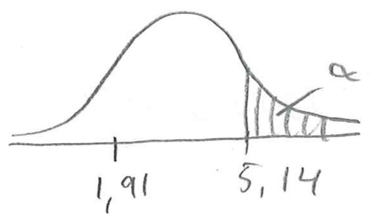
Egenkontroll $A > B > C$

Bygg ANOVA-tabell

Variationskälla	SS	df	MS	F_0
mellan prod. enhet	$B-C =$ 46 667	2	$SS_B/2 =$ 23 333	$\frac{MS_B}{MS_E} =$ 1,91
Okänd	$A-B =$ 76 333	6	$SS_E/6 =$ 12 222	
Totalt	$A-C =$ 120 000	8		

⑤ Jämför F_0 med F från tabellen för att se om vi kan förkasta H_0

$F_0 = 1,91 < F = 5,14$.



Vi kan ej förkasta H_0, att genomsnittslivslängden skiljer mellan olika produktionsenheter.

Uppgift 3.8 Tre olika investeringsrådgivare gav en bedömning av den riskjusterade utvecklingen av sex finansiella fonder.
Kan utifrån datan se indikatorer på om de tre rådgivarna gav väsentligt skilda bedömningar?

Fond	I	II	III
1	81	76	70
2	51	51	43
3	67	70	68
4	88	75	83
5	41	57	45
6	59	61	60

Tvåvägs variansanalys

1. $H_0: \mu_I = \mu_{II} = \mu_{III}$
   $H_1:$ alla ej lika
2. $\alpha = 0,05$
3. Testvariabel $F = \dfrac{MS_{rådgivare}}{MS_E}$
4. $A = \sum_{i=1}^{6} \cdot \sum_{j=1}^{3} \cdot x_{ij}^2 = 81^2 + 51^2 \ldots = 74\,456$

$B = \sum_{i=1}^{3} \cdot \left(\sum_{j=1}^{6} \cdot x_{ij}\right)^2 = 70\,471,7 = \frac{1}{6}(81+51..)^2$

$C = \sum_{i=1}^{6} \left(\sum_{j=1}^{3} \cdot x_{ij}\right)^2 = \frac{1}{3}(81+76+70 \ldots)^2 = 74\,208$

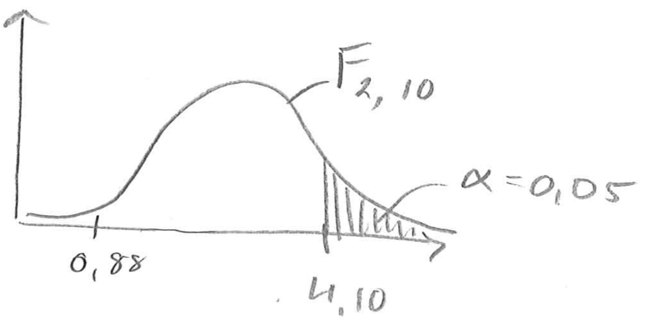
$D = \left(\sum_{i=1}^{6} \cdot \sum_{j=1}^{3}{}^{n_j} \cdot x_{ij}\right)^2 / N = \frac{1}{18}(81+51 \ldots)^2 = 70\,437,6$

A > D B > D C > D

ANOVA - tabell

Variationskälla	SS	df	MS	F_0
mellan rådgivare	B − D = 34,11	2	17,05	0,880
mellan fonder	C − D = 3770,4	5	754,08	38,39
okänd	$SS_T - (SS_R + SS_F) =$ 192,9	10	19,39	
totalt	A − D = 3998,4	17		

$$F_0 = \frac{17,05}{19,39} = \frac{\text{variation}}{\text{fel}} \quad \text{och} \quad \frac{754,08}{19,39} = 38,39$$



5. F_0 rådgivare $< F_{2,10}^{0,05}$ => Vi kan ej förkasta H_0

$0,88 < 4,10$

121210

Exempel och tentor

3,15

		Temperatur		
		100°C	200°C	300°C
Tid	5 h	2, 3	7, 10	16, 15
	10 h	14.16	16.19	18.20

a) Förkommer det någon samspeleffekt mellan tid och temperatur?

Variationskälla	SS	df	MS	F
Temperatur	144,5	2		
Tid	208,33	1		
Samspel	41.167	2	20,58	
Okänd	14,00	6	2,33	
Totalt		11		

← Alltid 2 df på samspel

H_0 : Inget samspel

H_1 : det finns ett samspel

$\alpha = 0,05$

$$F_0 = \frac{MS_{samspel}}{MS_{okänd}} = 8,33$$

Jämför F_0 med $F^{\alpha=0,05}_{\nu_1=2 \text{ och } \nu_2=6} \Rightarrow 5,14$

Vi förkastar H_0! Det finns samspel mellan tid och temp.

b) Skillnad mellan tider?

Steg 1. $H_0: \mu_1 = \mu_2$

H_1: alla ej lika $\mu_1 \neq \mu_2$

2. $\alpha = 0,05$

3. $\frac{MS_{tider}}{MS_{okänd}} = \frac{208,33}{2,33} = 89,3$

4. ANOVA tabell

5. Jämför F_0 med $F^{\alpha=0,05}_{r_1=1, v_2=6} = 5,9$

Förkasta H_0! Det finns skillnad mellan tiderna.

c) Samband mellan temperaturen?

1. $H_0: \mu_1 = \mu_2 = \mu_3$

H_1: alla är ej lika

2. $\alpha = 0,05$

3. $F_0 = \frac{MS_{temp}}{MS_{okänd}} = \frac{72,2}{2,33} = 30,97$

4. ANOVA-tabell

5. Jämför F_0 med $F^{\alpha=0,05}_{r_1=2, v_2=6} = 5,14$

Förkasta H_0! Det finns ett samband

Övningstenta 1

1. Bill och Georg går till puben. De beslutar sig för att spela dart. Sedan gammalt vet de att Bill träffar tavlan med sannolikhet 0,7 medan Georg oberoende av Bills resultat träffar tavlan med sannolikhet 0,4. De båda vännerna kastar varsin pil mot tavlan samtidigt.

a) Antag att endast en pil träffar tavlan. Vad är sannolikheten att det är Georg som har träffat den?

$P(\text{Bill}) = 0,7 \qquad P(\text{Georg}) = 0,4 \qquad$ Oberoende

$$P(\text{Georg} \mid \text{Endast en pil}) = ? \Rightarrow \frac{P(\text{Georg} \cap \text{endast en})}{P(\text{endast en})} =$$

$$= \frac{P(G \cap B^c)}{P(G \cap B^c) + P(G^c + B)} \overset{\text{oberoende}}{=} \frac{P(G) \cdot P(B^c)}{P(G) \cdot P(B^c) + P(G^c) \cdot P(B)} =$$

$$= \frac{0,4 \cdot 0,3}{04 \cdot 0 \cdot 3 + 0,6 \cdot 0,7} = 0,22\ldots$$

b) Antag att tavlan träffas av minst en pil. Vad är sannolikheten att Georgs pil har träffat?

$$P(G \mid \text{minst en pil}) = \frac{P(G \cap \text{minst en pil})}{P(\text{minst en pil})} = \frac{P(G)}{1 - P(\text{ingen pil})} =$$

$$= \frac{P(G)}{1 - P(G^c) \cdot P(B^c)} = \frac{0,4}{1 - (0,6 \cdot 0,3)} = 0,4878$$

(2.) Antag att man har 2 stokatiska variabler ξ och η.
Sannolikhetsfördelningen för ξ

$\xi = x$	$P(\xi = x)$
0	0,1
1	0,1
2	0,2
3	0,1
4	0,3
5	0,2

Den stokastiska variabeln η beräknas mha sambandet $\eta = (\xi - 2)^2$. Beräkna väntevärdet och variansen för η.

$E(\eta) = \sum_{i \in D} x_i \cdot p(x_i)$

$= 0 \cdot 0,2 + 1 \cdot 0,2 + 4 \cdot 0,4 + 9 \cdot 0,2 = \underline{3,6}$

$\eta = y$	$P(\eta = y)$
0	0,2
1	0,1+0,1=0,2
4	0,1+0,3=0,4
9	0,2

$P(\eta = y)$ räknas ut när $P(\eta = y) = 0$ ex alltså $(\xi - 2)^2 = 0$, när $\xi = 2 \Rightarrow P = 0,2$

$Var(\eta) = E[\eta - \mu]^2 =$

$E(\eta^2) - E(\eta)^2 = \sum_{i \in D} x_i^2 \cdot P(x_i) - E(x_i)^2$

$0^2 \cdot 0,2 + 1^2 \cdot 0,2 + 4^2 \cdot 0,4 + 9^2 \cdot 0,2 - 3,6^2 = 9,84$

③ Låt ξ beteckna avståndet i m som ett djur måste förflytta sig från sin födelse tills det hittar ett ledigt revir. För en speciell pungråtta gäller att ξ är exponentialfördelad med $\lambda = 0,01386$.

a) Beräkna sannolikheten att råttan behöver flytta sig mer än 100 m.

$$P(\xi > 100) = 1 - P(\xi < 100) = 1 - (1 - e^{-100 \cdot 0,01386}) = 0,25$$

b) Antag att man studerar 10 slumpmässigt valda råttor. Vad är sannolikheten att minst en av de måste flytta sig längre än 100 m?

η = antal råttor som rör sig längre än 100 m

$\eta \sim Bin(n, p) = Bin(10, 0,25)$

$$P(\eta \geq 1) = 1 - P(\eta < 1) = 1 - \binom{10}{0} \cdot 0,25^0 \cdot (1 - 0,25)^{10} = 0,9437$$

Exempel från tenta 19 dec 2007

(1.) Man skiljer på 4 olika blodtyper, 0, A, B, AB. Andelen som har varje blodtyp är

0 44%
A 42%
B 10%
AB 4%

a) Antag man slumpmässigt väljer 2 pers. Vad är sannolikheten att båda har A?

$$P(\text{person } 1\, A) \cdot P(\text{person } 2\, A) = P(A) \cdot P(A) = 0,42 \cdot 0,42 = 0,1764$$

b) Vad är sannolikheten att de har samma blodtyp?

4 utfall: Båda A, Båda B, Båda 0, Båda AB =

$$P(A) \cdot P(A) + P(B) \cdot P(B) + P(0) \cdot P(0) + P(AB) \cdot P(AB) = 0,42^2 + 0,1^2 + 0,44^2 + 0,04^2 = 0,3816$$

4. forts. ÖT

I ett fullständigt 2^3-faktorförsök ville man undersöka effekten av 3 faktorer A, B, C. För att kunna göra en bättre analys kompletterade man med 3 försök i centrumpunkten.

Försöksordning	A	B	C	resultat
7	−	−	−	−4
3	+	−	−	9
1	−	+	−	−1
6	+	+	−	16
4	−	−	+	−1
5	+	−	+	13
2	−	+	+	2
8	+	+	+	21
9	M	M	M	7
10	M	M	M	4
11	M	M	M	3

$l_A = 13{,}75$, $l_B = 5{,}25$, $l_C = 3{,}75$, $l_{AB} = 2{,}25$ $l_{AC} = 0{,}75$
$l_{BC} = 0{,}25$ $l_{ABC} = 0{,}25$

a) Beräkna ett referensintervall och avgör vilka av effekterna som är signifikanta.

Vi utnyttjar centrumpunktsmätningarna för att uppskatta variansen. $s^2 = \dfrac{\sum_i y_i^2 - \dfrac{(\sum_i y_i)^2}{n}}{n-1} = \dfrac{7^2+4^2+3^2-\dfrac{(7+4+3)^2}{3}}{2}$

$s^2 = 4{,}33$ med 2 df

Ett 95% konfidensintervall för effekten ges av följande:

$$0 \pm t_{2,\,0{,}05} \cdot \frac{2s}{\sqrt{N}} = 0 \pm 4{,}303 \cdot \frac{2 \cdot \sqrt{4{,}33}}{\sqrt{8}}$$

$N = 8$ st mätningar utan centrumpunkterna

$= 0 \pm 6{,}33 \Rightarrow$ Endast A är stor nog och är den enda som är signifikant, de andra ger "ingen" effekt

$l_A = 13{,}75 > 0 \pm 6{,}33$

b) Beskriv försöksresultatet med en matematisk formel.

Det är bara A som är signifikant

$\hat{y} = l_M + \frac{l_A}{2} \cdot x_A$, där l_M beräknas på de 8 punkter som ej är centrumpunkter

$l_M = \frac{1}{8}(-4+9-1-16+1+13+2+21) = 6{,}875$

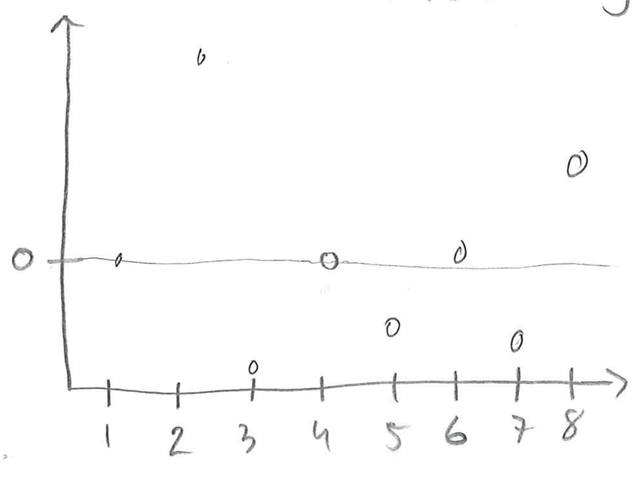
$\hat{y} = 6{,}875 + \frac{15{,}75}{2} \cdot x_A$ $\quad x_A$ = Nivå på A, låg $= -1$, hög $= +1$

Centrumpunkter är medelvärdet på de olika nivåerna, ex $A+ = 20$ $A- = 10$ $M = 15$

c) Genomför en residualanalys och avgör om de antaganden man gör på försöksresultaten är uppfyllda.

Plotta i försöksordningen skillnaden $\varepsilon = y - \hat{y}$ (residualen)

Felet mellan verklighet och modell



6. Ett reducerat 2^{6-3}-försök skall genomföras.

a) Ange generatorer (de du tror är lämpliga, de behöver inte vara de bästa)

$A \quad B \quad C \mid D, E \quad F$

$D = AB \quad E = AC \quad F = BC$

b) Vilken upplösning får du?

$I_1 = D \cdot D = D \cdot AB = ABD$

$I_2 = E \cdot E = E \cdot AC = ACE$

$I_3 = F \cdot F = F \cdot BC = BCF$

$I_4 = I_1 \cdot I_2 = \cancel{A}BD \cdot \cancel{A}CE = BCDE$

$I_5 = I_1 \cdot I_3 = A\cancel{B}D \cdot \cancel{B}CF = ACDF$

$I_6 = I_2 \cdot I_3 = A\cancel{C}E \cdot B\cancel{C}F = ABEF$

$I_7 = I_1 \cdot I_2 \cdot I_3 = \cancel{A}\cancel{B}D \cdot \cancel{A}\cancel{C}E \cdot \cancel{B}\cancel{C}F = DEF$

Upplösning = minsta antal faktorer i försöken = $\underline{\overline{III}}$

c) Ange alias som faktorn B får.

$l_B = B + AD + ABCE + CF + CDE + ABCDF + AEF + BDEF$

$BI_1 = B \cdot ABD = AD$

$BI_2 = B \cdot ACE = ABCE$

$BI_3 = B \cdot BCF = CF$

$BI_4 = B \cdot BCDE = CDE$

$BI_5 = B \cdot ACDF = ABCDF$

$BI_6 = B \cdot ABEF = AEF$

$BI_7 = B \cdot DEF = BDEF$

7. Vid upprepade mätningar av ett föremål fick man en genomsnittsvikt på $\bar{x} = 15,10$. Vågen ger ett slumpmässigt fel som är normalfördelat med $\mu = 0$ och $\sigma = 0,10$ g. Anta att man har gjort 25 mätningar.

a) Ange ett 95%-igt konfidensintervall för föremålets verkliga vikt.

Ett 95%-igt konfidensintervall ges av $\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$15,10 \pm 1,96 \cdot \frac{0,10}{\sqrt{25}} \Rightarrow 15,10 \pm 0,0392$$

b) Testa på 5% signifikansnivå om den sanna genomsnittsvikten kan vara större än 15,05 g

1. $H_0 : \mu \leq 15,05$

   $H_1 : \mu > 15,05$

2. $\alpha = 0,05$

3. Normalfördelad data med känt σ

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

4. Urvalet gav $Z = \frac{15.10 - 15.05}{0.10/\sqrt{25}} = 2,5$

5. Jämför Z med $Z_\alpha = 1,645$

$$Z = 2,5 < Z_\alpha = 1,645$$

(8.) En plastfabrikant vill undersöka om 3 olika kemiska ingredienser A, B, C har olika genomsnittseffekt på elasticiteten hos en produkt.

Produkt	Resultat
A	5 6 5 8 6 7 6 5 6 7
B	8 9 8 7 9 9 10 8
C	10 10 9 8 8 9 10 9 8 9 10 8

Följande kvadratsummor beräknas
produkt : 49,8
Totalt = 72,7

ANOVA

(1) $H_0: \mu_A = \mu_B = \mu_C$
H_1: Alla ej lika

(2) $\alpha = 0,05$

(3) $F = \frac{MS_{produkt}}{MS_{okänd}}$

(4)

Variationskälla	SS	df	MS	F_0
Produkt	49,8	2	24,9	29,36
Okänd	22,9	27	0,8481	
Totalt	72,7	29		

(5.) Jämför F_0 med $F^{\alpha=0,05}_{r_1=2,\, r_2=27}$ $F_0 > F_1$

Förkasta H_0, olika ingredienser ger olika effekt

$SS_{Produkt} = B - C \qquad SS_{okänd} = A - B \qquad SS_T = A - C$

$$A = \sum_{j=1}^{k} \cdot \sum_{i=1}^{n_j} x_{ij}^2 = 5^2 + 6^2 + 5^2 + \ldots + 9^2 + 10^2 + 8^2 =$$

$$B = \text{se tabell} = \frac{1}{3}(5+6+5)^2 + \ldots + (9+10+8)^2 =$$

$$C = \text{-- " --} = \frac{(5+6\,5+\ldots+9+10+8)^2}{30}$$