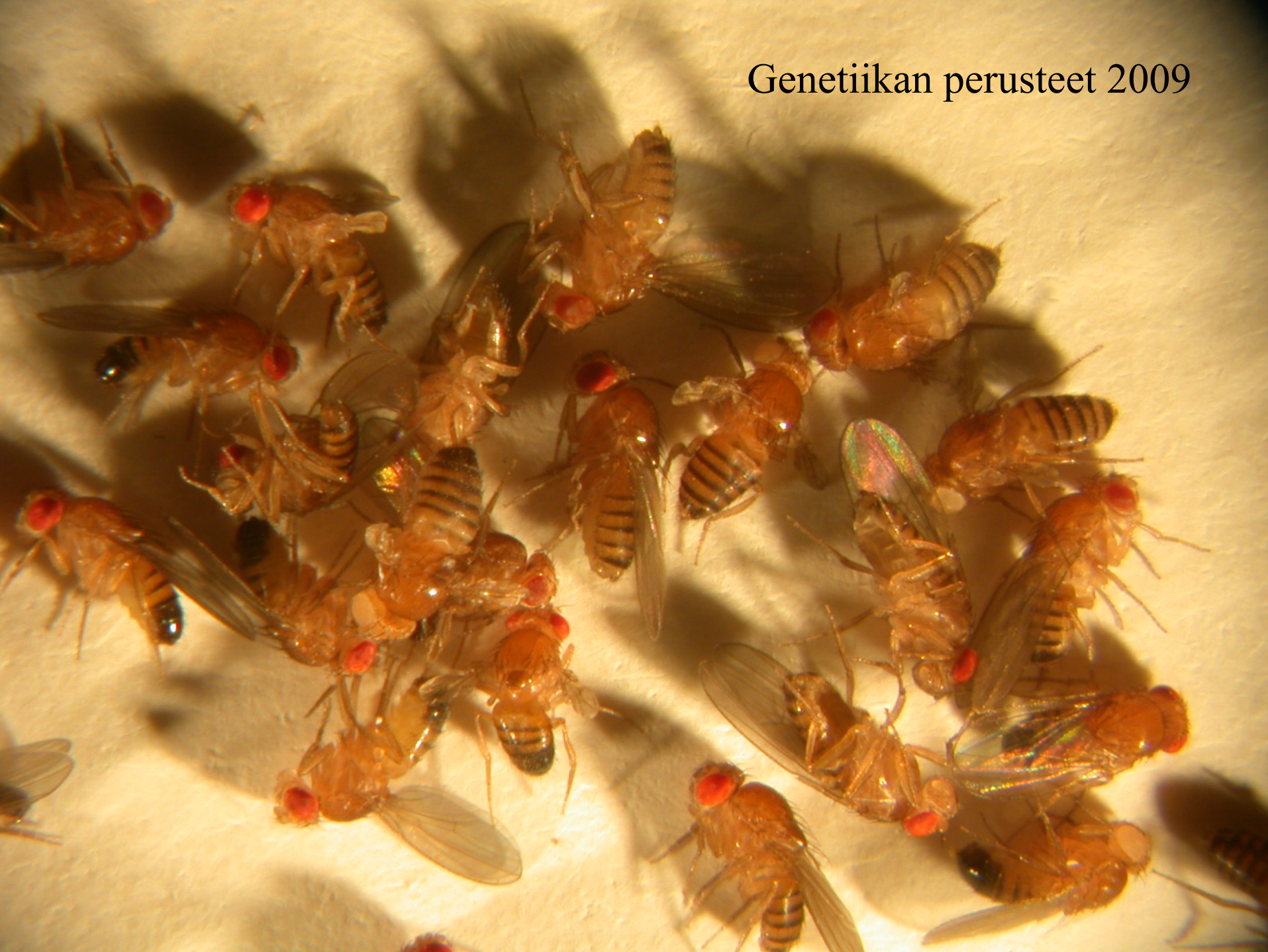




Malli **selittää**, mutta myös **ennustaa** ja ennusteen voi **testata** kokeella.

Mendel testasi F_2 -mallinsa tuottamalla itsepölytyksellä F_3 -polven

Seuraava sukupolvi tai toinen, riippumaton koe ovat voimakkaita testejä. Saman ilmiön havaitseminen muilla eliöillä ylentää hypoteesin teoriaksi

Malli **ennustaa** ja ennusteen voi **testata** kokeella.

Yksikin koe voidaan testata ***tilastollisesti***

Muodostetaan ***nollahypoteesi*** mallin mukaan

Tutkitaan, poikkeako havaintojoukko nollahypoteesista epätavallisen paljon vai ei

Esimerkki testaamisesta:

Tynkäsiipisiä punasilmäisiä kärpäsiä (vg ; $+/vg$: $+$) on risteytetty normaalisiipisten ruskeasilmäisten kärpästen kanssa ($+$; $se/+$; se) ja F_1 -jälkeläiset ovat ristetyneet vapaasti keskenään.

Eri työparit ovat sitten laskeneet F_2 -polven kärpäsensä niin hyvin kuin ovat osanneet.

Testataan, noudattaako "yhteensä"-sarakkeen tulos mallia 9:3:3:1, joka sopii käytettäväksi, koska ensimmäisen polven heterotsygootit olivat molempien ominaisuuksien suhteen villityyppiä (siis $+$ -alleeli dominoiva).

Fenotyyppi	Katri & Tuula	Mirja & Ulla	Kaisa & Katriina	Sirpa & Marja	Riikka & Teri	Katja & Maarit	Jesse & Joni	Janne & Minna	yhteensä <i>o</i>	odotusarvo <i>e</i>
+ ; +	91	115	221	61	213	199	87	192	1179	1068.75
vg; +	34	15	40	4	76	68	28	42	307	356.25
+ ; <i>se</i>	14	31	58	14	69	65	25	37	313	356.25
vg; <i>se</i>	4	2	21	3	20	26	5	20	101	118.75
Σ	143	163	340	82	378	358	145	291	1900	
X²	12.2	20.4	14.0	14.2	1.0	0.7	2.4	13.6		26.09

Fenotyyppien **odotusarvot** (*e*) yllä olevan taulukon "yhteensä" sarakkeessa ovat 9:3:3:1 mallin perusteella

$$(9/16) \times 1900 = 1068.75$$

$$(3/16) \times 1900 = 356.25$$

$$(3/16) \times 1900 = 356.25$$

$$(1/16) \times 1900 = 118.75$$

Testinä käytetään χ^2 -testiä. Lasketaan *testisuure*

$$X^2 = \sum [(o-e)^2/e] \quad \text{tai}$$

$$G^2 = \sum [2 \cdot o \cdot \ln(o/e)]$$

Sitten katsotaan χ^2 -taulukosta raja-arvot.

Luokkia on neljä, joten ns. vapausasteita on kolme.

Saamamme X^2 testisuure **26.09** on paljon suurempi kuin raja-arvo 16.266, jota suuremman mallia noudattava jakautuma sattumalta saa harvemmin kuin kerran tuhannesta ($P < 0.001$).

Taulukko χ^2 . Kriittisiä arvoja χ^2 -jakaumaan perustuvissa testeissä. Oikean vertailuarvon löytämiseksi valitaan ensin vapausasteluku taulukon reunasta ja sen jälkeen vertailuarvo halutun riskitason kohdalta. Esimerkiksi vapausasteilla $df = 5$ ja riskitasolla $\alpha = 0.05$ on kriittinen arvo 11.070.

α					α				
df	0.1	0.05	0.01	0.001	df	0.1	0.05	0.01	0.001
1	2.706	3.841	6.635	10.828	51	64.295	68.669	77.386	87.968
2	4.605	5.991	9.210	13.816	52	65.422	69.832	78.616	89.272
3	6.251	7.815	11.345	16.266	53	66.548	70.993	79.843	90.573
4	7.779	9.488	13.277	18.467	54	67.673	72.153	81.069	91.872
5	9.236	11.070	15.086	20.515	55	68.796	73.311	82.292	93.168
6	10.645	12.592	16.812	22.458	56	69.918	74.468	83.513	94.460
7	12.017	14.067	18.475	24.322	57	71.040	75.624	84.733	95.751
8	13.362	15.507	20.090	26.124	58	72.160	76.778	85.950	97.039
9	14.684	16.919	21.666	27.877	59	73.279	77.931	87.166	98.324
10	15.987	18.307	23.209	29.588	60	74.397	79.082	88.379	99.601

riskitaso

Fenotyyppi	Katri & Tuula	Mirja & Ulla	Kaisa & Katriina	Sirpa & Marja	Riikka & Teri	Katja & Maarit	Jesse & Joni	Janne & Minna	yhteensä <i>o</i>	odotusarvo <i>e</i>
+ ; +	91	115	221	61	213	199	87	192	1179	1068.75
vg; +	34	15	40	4	76	68	28	42	307	356.25
+ ; <i>se</i>	14	31	58	14	69	65	25	37	313	356.25
vg; <i>se</i>	4	2	21	3	20	26	5	20	101	118.75
Σ	143	163	340	82	378	358	145	291	1900	
χ^2	12.2	20.4	14.0	14.2	1.0	0.7	2.4	13.6		26.09

Testi siis osoitti, että tulos poikkeaa “erittäin merkitsevästi” 9:3:3:1 mallin mukaisesta ennusteesta (nollahypoteesista) ($P < 0.001$)

Tarkoittaako se sitä, että Mendel olisikin väärässä?

No ei. Se voi tarkoittaa sitä, että mutanttien elinkyky on heikompi kuin heterotsygoottien ja villien homotsygoottien elinkyky

Fenotyyppi	Katri & Tuula	Mirja & Ulla	Kaisa & Katriina	Sirpa & Marja	Riikka & Teri	Katja & Maarit	Jesse & Joni	Janne & Minna	yhteensä <i>o</i>	odotusarvo <i>e</i>
+ ; +	91	115	221	61	213	199	87	192	1179	1068.75
vg; +	34	15	40	4	76	68	28	42	307	356.25
+ ; <i>se</i>	14	31	58	14	69	65	25	37	313	356.25
vg; <i>se</i>	4	2	21	3	20	26	5	20	101	118.75
Σ	143	163	340	82	378	358	145	291	1900	
X^2	12.2	20.4	14.0	14.2	1.0	0.7	2.4	13.6		26.09

Jokaisen työparin tulos on testattu myös erikseen.

Joillakin tulos poikkeaa odotetusta hyvin vähän (odotusarvoja ei ole esitetty), joillakin paljon. Pullot ovat erilaisia, saalis vaihtelee

Jos kaikkien poikkeama olisi samansuuntainen, olisi testisuureiden summa (78.5) = summataulukon testisuure (X^2 on additiivinen)

Fenotyyppi	Katri & Tuula	Mirja & Ulla	Kaisa & Katriina	Sirpa & Marja	Riikka & Teri	Katja & Maarit	Jesse & Joni	Janne & Minna	yhteensä <i>o</i>	odotusarvo <i>e</i>
+ ; +	91	115	221	61	213	199	87	192	1179	1068.75
vg; +	34	15	40	4	76	68	28	42	307	356.25
+ ; <i>se</i>	14	31	58	14	69	65	25	37	313	356.25
vg; <i>se</i>	4	2	21	3	20	26	5	20	101	118.75
Σ	143	163	340	82	378	358	145	291	1900	
X^2	12.2	20.4	14.0	14.2	1.0	0.7	2.4	13.6		26.09

Jos kaikkien poikkeama olisi samansuuntainen, olisi testisuureiden summa (78.5) = summataulukon testisuure (X^2 on *additiivinen*)

Erotus $78.5 - 26.1 = 52.4$ on suuri, ja käy homogeenisuustestistä: työparien tulokset eroavat merkitsevästi toisistaan

Selitys: kärpäset liiskautuvat poropeukaloilla helpommin

χ^2 -jakautuman ja testin idea.

Jos meillä on *äärellinen* näyte esimerkiksi kärpäsiä, jotka noudattavat uskollisesti 9:3:3:1 mallia, eivät todelliset luvut tietenkään ole ihan täsmälleen odotusarvojen mukaiset, vaan pelkkä satunnaisvaihtelu sysii niitä hiukan milloin mihinkin suuntaan.

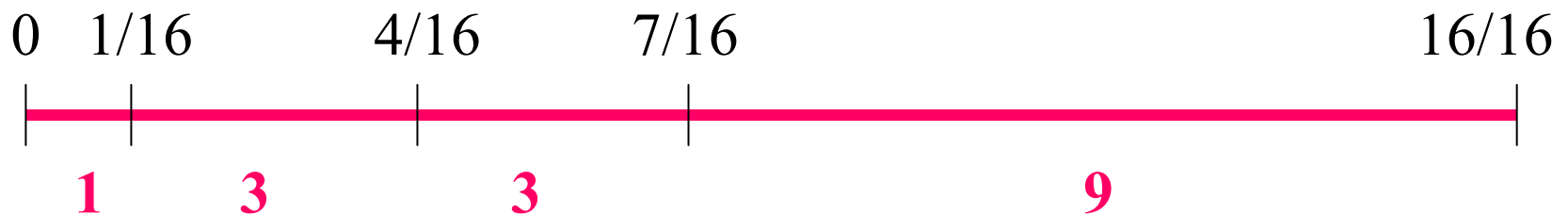
Pienet poikkeamat eivät vielä vakuuta ketään siitä että jotakin olisi pielessä, eli että malli (nollahypoteesi) ei pitäisi paikkaansa (ko. tapauksessa).

Kuinka suuret poikkeamat sitten saavat meidät ajattelemaan, että pelkän 9:3:3:1 mallin lisäksi jokin tekijä vieroittaisi havaittua tulosta odotetusta? Tutkitaanpa asiaa **simulaatiolla**.

Simulaation periaate, esimerkkinä dihybridisteytys

1) arvotaan satunnaisluku RND väliltä 0-1

2) tutkitaan sen sijoittuminen lukusuoralle:



3) "tehdään" virtuaalifenotyyppi":

jos $RND < 1/16$ niin "virtuaalifenotyyppi" on **ab**

jos $1/16 < RND < 4/16$ niin se on **Ab**

jos $4/16 < RND < 7/16$ niin se on **aB**

jos $RND > 7/16$ niin virtuaalifenotyyppi on **AB**

4) toistetaan, kunnes on saatu haluttu määrä yksilöitä

RANDOMIZE TIMER

10

FOR i = 1 TO 100

a = RND

IF a < (1 / 16) THEN p1 = p1 + 1: GOTO 100

IF a < (4 / 16) THEN p2 = p2 + 1: GOTO 100

IF a < (7 / 16) THEN p3 = p3 + 1: GOTO 100

p4 = p4 + 1

100

NEXT i

PRINT p1, p2, p3, p4

p1 = 0: p2 = 0: p3 = 0: p4 = 0

GOTO 10

Finished - Microsoft QuickBASIC

Auto

5	19	15	61
7	20	17	56
7	23	16	54
4	13	31	52
9	16	18	57
8	20	23	49
6	18	21	55
6	14	19	61
5	21	14	60
8	26	10	56
5	14	20	61
6	21	15	58
8	20	14	58
9	13	19	59
6	21	20	53
9	15	19	57
8	19	19	54
9	17	14	60
8	18	10	64
5	15	23	57
10	19	19	52
6	19	23	52
3	19	23	55

ab	Ab	aB	AB	
6.25	18.75	18.75	56.25	odotettu/100

Lista: *fenotyyppien* lukumääriä (n= 100) erillisistä arpajaisista

Simulointiohjelma

Dihybridisimulaatio

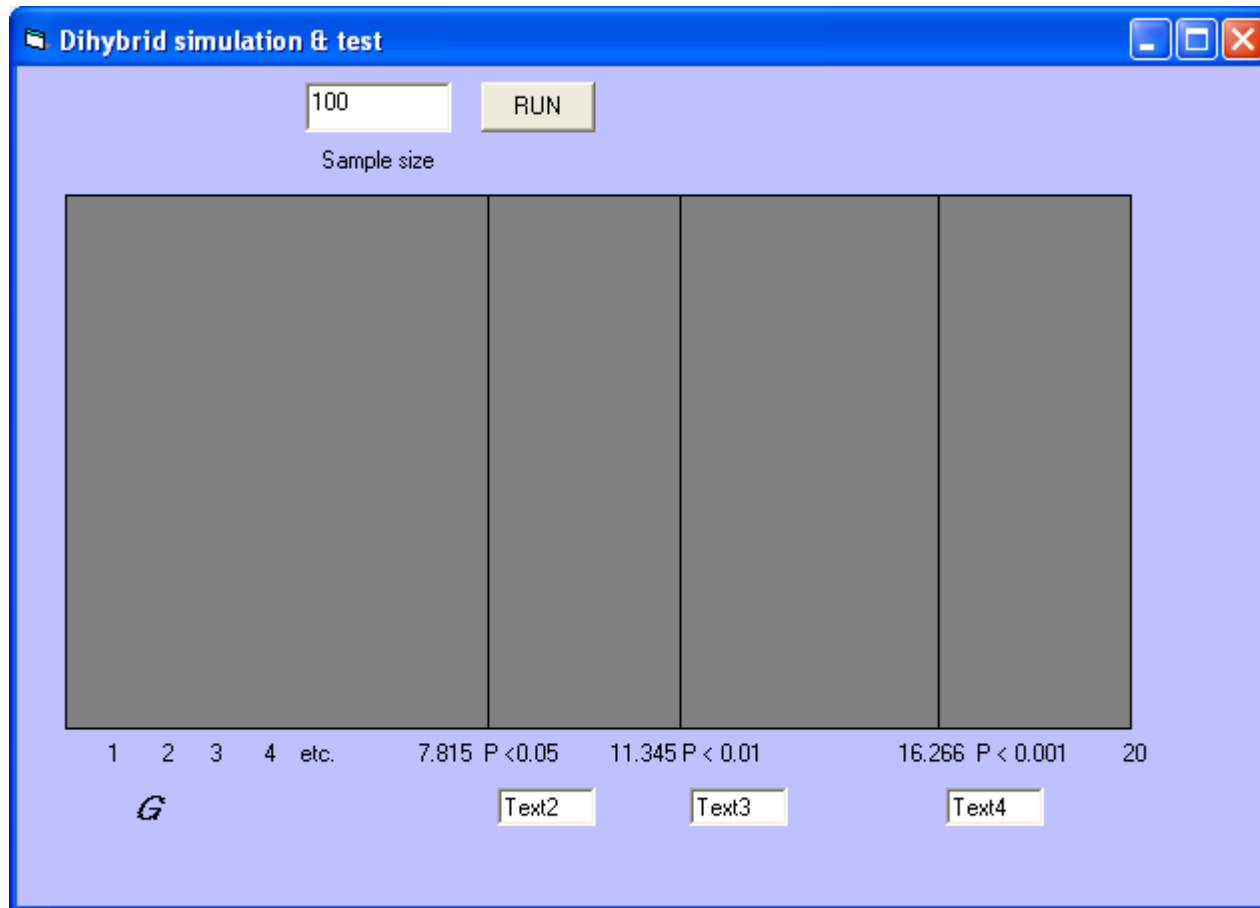
arpoo dihybridiristeytystuloksia 9:3:3:1 -**mallin mukaan**
mutta **satunnaisesti** ja laskee sitten testisuureen

G^2 sekä vertaa sitä χ^2 -taulukon raja-arvoihin

Menettelyä sanotaan (**stokastiseksi**) **simulaatioksi**

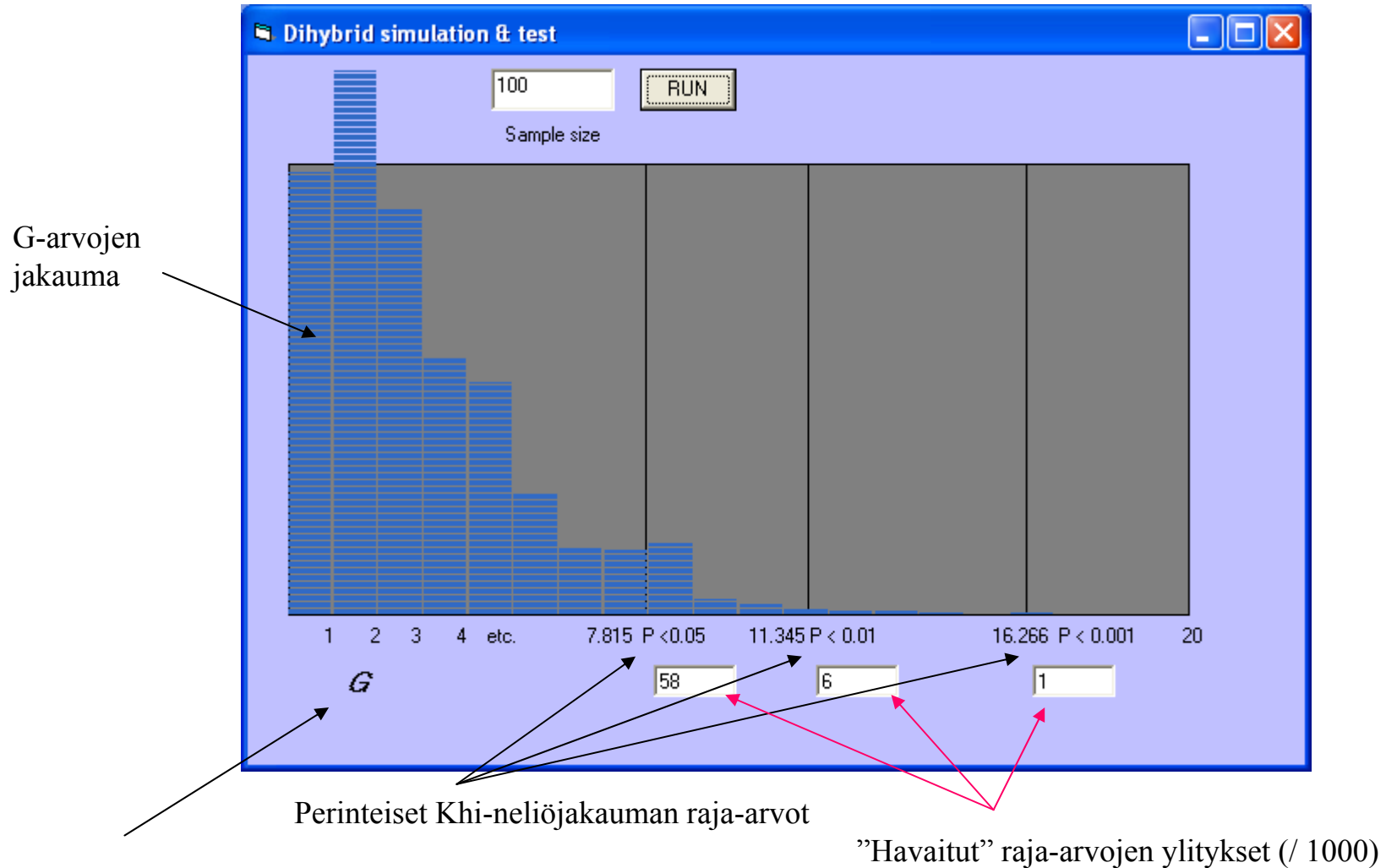
$$G^2 = \Sigma[2 \cdot o \cdot \ln(o/e)]$$

Dihybridisimulaatio



Dihybridiristeytyksen testin testilomake

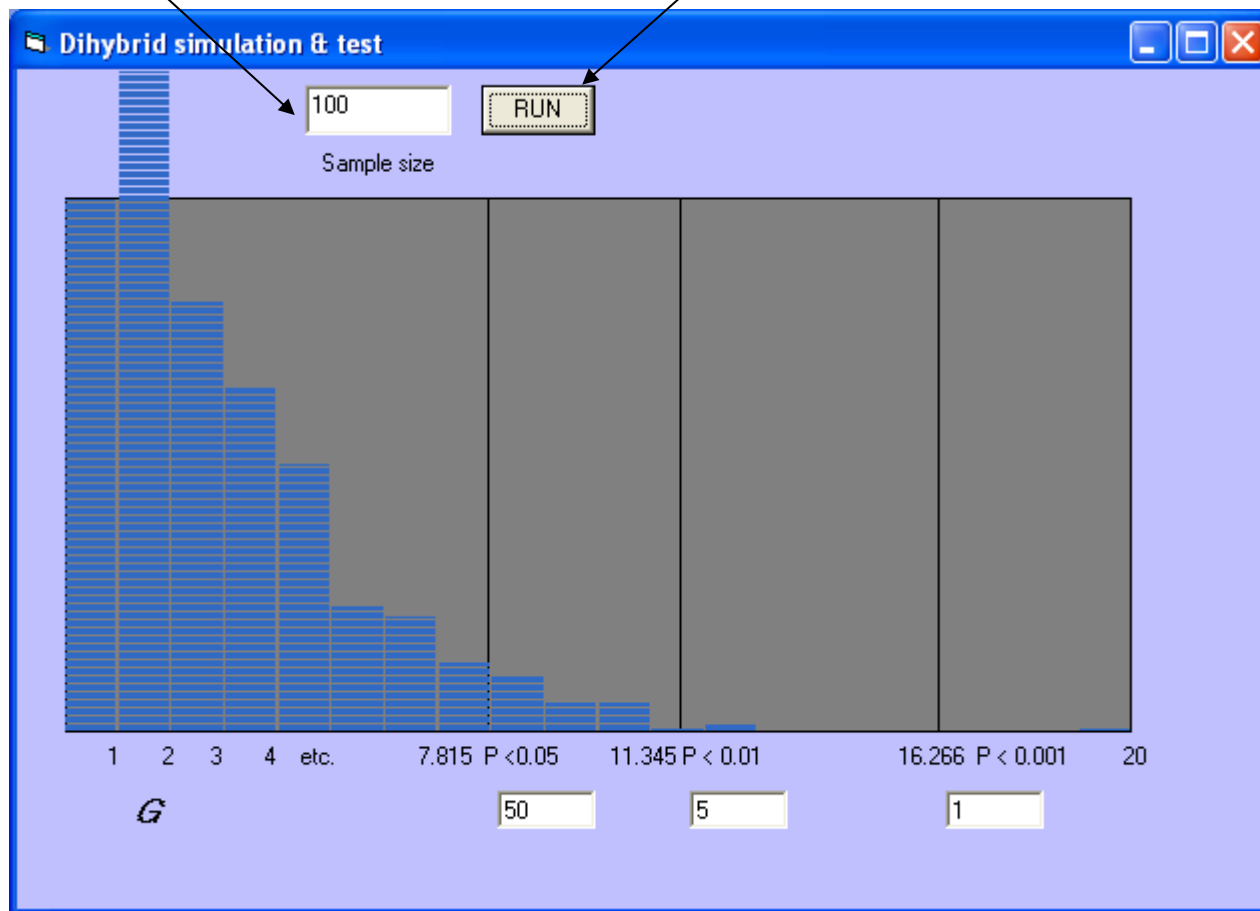
Tuhat virtuaalikoetta (ja vajaata sekuntia) myöhemmin:

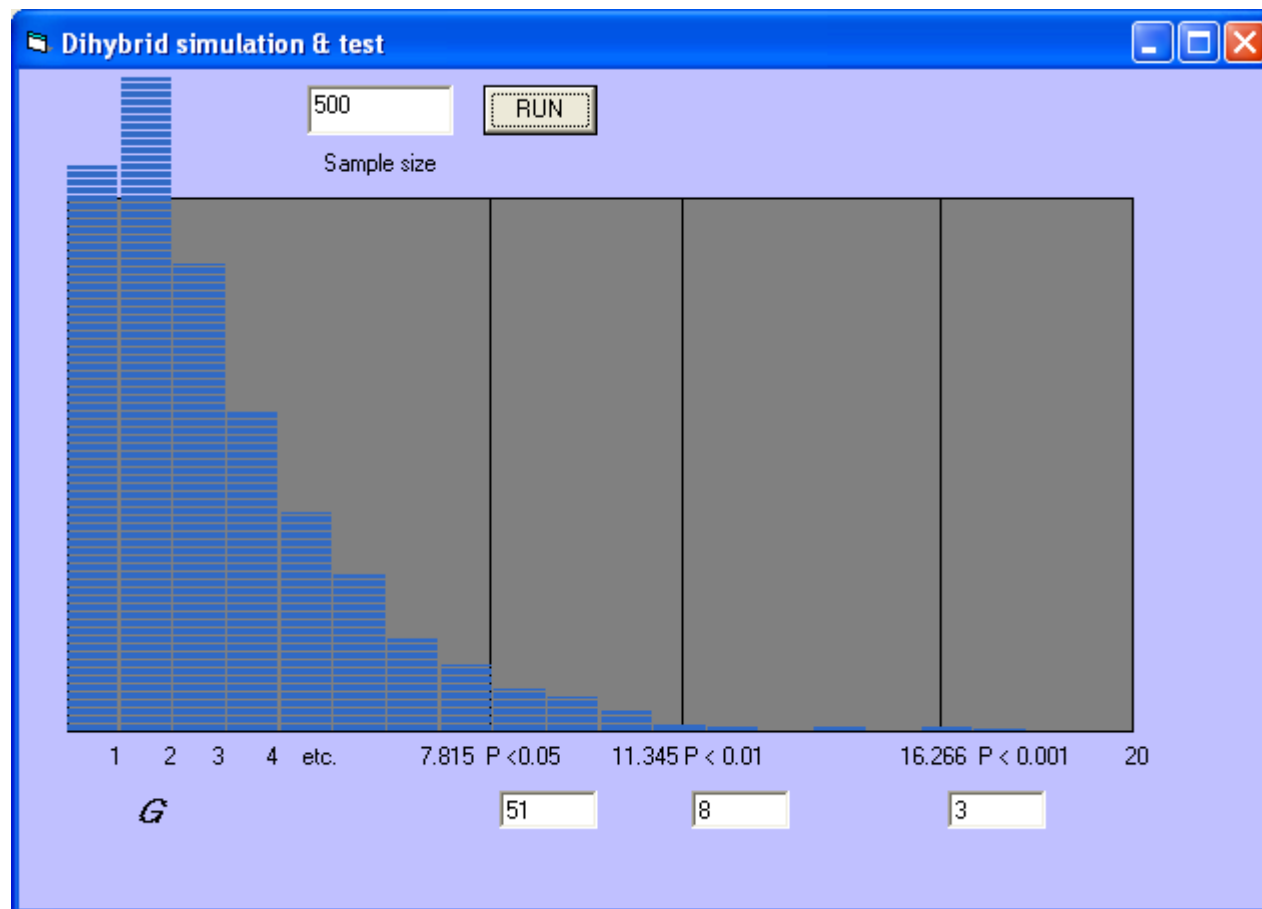


$$G^2 = \Sigma[2 \cdot o \cdot \ln(o/e)]$$

Voit tutkia näyttekoon vaikutusta testitulokseen. Juu onkin siinä, että testi standardoi erikokoiset kokeet samaan asteikkoon!

Ajonappula pyyhkii myös taulun





Dihybridisimulaatio

