SF1901: Sannolikhetslära och statistik
Föreläsning 11.
Mer om konfidensintervall.
Hypotesprövning

Jan Grandell & Timo Koski

22.02.2016

KTH Matematik

Innehåll

- Konfidensintervall för skillnaden mellan två väntevärden
  - Två stickprov
    - konfidensintervall för skillnaden mellan två väntevärden $\mu_1 - \mu_2$ i $N(\mu_1, \sigma_1)$ och $N(\mu_2, \sigma_2)$ med kända σ_1 och σ_2.
    - konfidensintervall för skillnaden mellan två väntevärdenmellan $\mu_1 - \mu_2$ i $N(\mu_1, \sigma_1)$ och $N(\mu_2, \sigma_2)$ med okända σ_1 och σ_2.
  - Stickprov i par/Matched Pairs
- Hypotesprövning: inledande exempel & Hypotesprövningens beståndsdelar

KTH Matematik

Sammanfattning (från tidigare)

Sats

Låt $x_1, \ldots, x_n$ vara ett slumpmässigt stickprov från $N(\mu, \sigma)$ där μ är okänt. Då är

$$I_\mu = (\bar{x} - \lambda_{\alpha/2} D, \bar{x} + \lambda_{\alpha/2} D)$$

om σ är känt $(D = \sigma/\sqrt{n})$

samt

$$I_\mu = \big(\bar{x} - t_{\alpha/2}(f)d, \bar{x} + t_{\alpha/2}(f)d\big)$$

om σ är okänt $(d = s/\sqrt{n},\ f = n-1)$

ett tvåsidigt konfidensintervall för μ med konfidensgraden $1 - \alpha$.

KTH Matematik

Två stickprov, konfidensintervall för skillnad mellan väntevärden.

Normalfördelning

Modell:

$$X_1, X_2, \ldots, X_{n_1} \quad \text{är} \quad N(\mu_1, \sigma_1) \quad \text{(stickprov 1)}$$
$$Y_1, Y_2, \ldots, Y_{n_2} \quad \text{är} \quad N(\mu_2, \sigma_2) \quad \text{(stickprov 2)}$$

där alla Xen och Yna är oberoende.

KTH Matematik

Två stickprov, konfidensintervall för skillnad mellan väntevärden.

a) σ_1 och σ_2 kända

Vi vill nu skaffa oss ett konfidensintervall för $\mu_1 - \mu_2$. En naturlig skattning av $\mu_1 - \mu_2$ är $\overline{X} - \overline{Y}$. Eftersom den är en linjärkombination av oberoende normalfördelade variabler, så gäller att

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

är $N(0,1)$-fördelad.

Två stickprov, konfidensintervall för skillnad mellan väntevärden.

Av detta leds vi till

$$I_{\mu_1-\mu_2} = \overline{x} - \overline{y} \pm \lambda_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,.$$

Om $\sigma_1 = \sigma_2 = \sigma$ reduceras detta till att

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

är $N(0,1)$-fördelad och

$$I_{\mu_1-\mu_2} = \overline{x} - \overline{y} \pm \lambda_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,.$$

KTH Matematik

Två stickprov, konfidensintervall för skillnad mellan väntevärden.

b) $\sigma_1 = \sigma_2 = \sigma$ okänd

Vi betraktar nu fallet då $\sigma_1 = \sigma_2 = \sigma$, men där σ är okänd. Detta skattas med s där s^2 är den *sammanvägda* stickprovsvariansen.
Man kan visa att man skall välja

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

och att

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

är $t(n_1 + n_2 - 2)$-fördelad.

KTH Matematik

Två stickprov, konfidensintervall för skillnad mellan väntevärden.

Vi får

$$\underline{I_{\mu_1-\mu_2} = \overline{x} - \overline{y} \pm t_{\alpha/2}(n_1+n_2-2)s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\,.$$

KTH Matematik

Sammanfattning (1)

Sats

Låt $x_1, \ldots, x_{n_1}$ *och* $y_1, \ldots, y_{n_2}$ *vara slumpmässiga, av varandra oberoende stickprov från* $N(\mu_1, \sigma_1)$ *respektive* $N(\mu_2, \sigma_2)$.
Om σ_1 *och* σ_2 *är kända så är*

$$I_{\mu_1-\mu_2} = (\bar{x} - \bar{y} - \lambda_{\alpha/2} D, \bar{x} - \bar{y} + \lambda_{\alpha/2} D)$$

ett tvåsidigt konfidensintervall för $\mu_1 - \mu_2$ *med konfidensgraden* $1 - \alpha$*; här är* $D = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

KTH Matematik

Sammanfattning

Sats

Om $\sigma_1 = \sigma_2 = \sigma$ *där* σ *är okänt så är*

$$I_{\mu_1-\mu_2} = (\bar{x} - \bar{y} - t_{\alpha/2}(f)d, \bar{x} - \bar{y} + t_{\alpha/2}(f)d)$$

ett tvåsidigt konfidensintervall för $\mu_1 - \mu_2$ *med konfidensgraden* $1 - \alpha$*; här är* $d = s\sqrt{1/n_1 + 1/n_2}$ *där* $f = (n_1 - 1) + (n_2 - 1)$.

KTH Matematik

Two samples = Två stickprov

In an experiment designed to test the effectiviness of paroxetine for treating bipolar depression, subjects were measured using the Hamilton Depression scale with results as follows:

Placebo group	$n_1 = 43$	$\overline{x} = 21.57, s_1 = 3.87$
Paroxenite treatment group	$n_2 = 33$	$\overline{y} = 20.38, s_2 = 3.91$

KTH Matematik

We assume that we have independent samples and normal distributions. We must assume this, we have just summaries of data, no chance of looking at histograms or boxplots.
We insert data in pooling (sammanvägning)

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(43-1)\cdot 3.87^2 + (33-1)\cdot 3.91^2}{43+33-2}$$

$$= 15.11$$

We choose $\alpha = 0.05$. The quantiles are found from the t-distribution with $43+33-2 = 74$ degrees of freedom. This gives us $t_{0.025} = 1.993$ (by >>tinv(0.975, 74) in Matlab).

Two samples; the confidence interval computed

$$s^2 = \frac{(43-1)\cdot 3.87^2 + (33-1)\cdot 3.91^2}{43+33-2} = 15.11$$

$$t_{0.025}(74) = 1.993$$

$$I_{\mu_1-\mu_2} = \overline{x} - \overline{y} \pm t_{\alpha/2}(n_1+n_2-2)s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= (21.57 - 20.38) \pm 1.993 \cdot \sqrt{15.11} \cdot \sqrt{\frac{1}{43}+\frac{1}{33}}$$

$$= 1.19 \pm 1.7930 = [-0.603, 2.938]$$

Två stickprov för paroxetine; Konfidensintervall med konfidens grad 0.95

$$I_{\mu_1-\mu_2} = [-0.603, 2.938]$$

Men vad säger detta för oss ? Vilken **kunskap** *om paroxetine har de givna data gett oss i form av detta konfidensintervall?*

Vi kommer att få ett svar med hjälp av teorin om hypotesprövning. Men först ett fall till.

KTH Matematik

Matched Pairs/Stickprov i par

We shall deal with a testing situation that is treacherously close to the comparison of two means (=två stickprov) above.
We present this by an example. Suppose that we want to test the effectiveness of a low-fat diet. The weight of n subjects is measured before and after the diet. The results are $x_1, \ldots, x_n$ *and* $y_1, \ldots, y_n$*, respectively. Obviously* x_i *and* y_i *would be dependent, but the samples corresponding to different subjects are independent.*

KTH Matematik

Matched Pairs

We have

Diet	Subject			
	1	2	...	n
weight before diet	x_1	x_2	...	x_n
weight after diet	y_1	y_2	...	y_n

KTH Matematik

Matched Pairs

Let us assume x_j for the jth subject is a sample from $N(\mu_j, \sigma_1)$ and y_j a sample from $N(\mu_j + \Delta, \sigma_2)$. Δ is the population mean difference for *all* matched pairs. Δ is the population parameter for the effectiveness of the low-fat diet.

KTH Matematik

Matched Pairs

- There is, as in case of two means, two series of observations. But the model for two means (=två stickprov) is inapplicable, the pairs x_j, y_j are now matched to each other, (two measurements of the weight of one and the same person). The data consists of n matched pairs.
- The unknown parameters are $\mu_1, \ldots, \mu_n$, σ_1, σ_2 och Δ.
- $\mu_1, \ldots, \mu_n$ reflect differences between subjects, whereas Δ reflects the systematic difference between the weights before and after the low fat diet. If $\Delta < 0$ then the weight after diet is in average lower than before the diet.
- Note that σ_1 and σ_2 can be different.

KTH Matematik

Stickprov i par

Det är lämpligt att utgå från ett praktiskt problem. Antag först att två personer A och B utför upprepade mätningar på *ett och samma objekt*. Om antalet mätningar är lika, får vi då mätvärden $x_1, \ldots, x_n$ respektive $y_1, \ldots, y_n$, vilka kan betraktas som två oberoende stickprov.

KTH Matematik

Stickprov i par

Om normalfördelning föreligger och personerna arbetar med samma (okända) precision, kan vi med den föregående analysen (två stickprov) avgöra t.ex. hur stor den systematiska skillnaden $\mu_1 - \mu_2$ är mellan personerna.

KTH Matematik

Stickprov i par

Nu antar vi i stället att varje person utför en mätning på vart och ett av n *olika* objekt. De erhållna mätvärdena kan skrivas

Person	Objekt			
	1	2	...	n
A	x_1	x_2	...	x_n
B	y_1	y_2	...	y_n

KTH Matematik

Stickprov i par

Det föreligger liksom förut två serier av mätvärden, men den tidigare modellen är oanvändbar, eftersom skillnader kan föreligga mellan objekten, alldeles oavsett om det föreligger skillnader mellan personerna eller ej. Tydligen hänger observationerna ihop på något sätt parvis, eftersom de två och två härrör från samma objekt.

KTH Matematik

Stickprov i par

En i många fall realistisk modell är följande. Värdet x_j för det j:te objektet antas komma från en normalfördelning $N(\mu_j, \sigma_1)$ och y_j från en annan fördelning $N(\mu_j + \Delta, \sigma_2)$.

KTH Matematik

Stickprov i par

Man har då okända parametrar $\mu_1, \ldots, \mu_n$, σ_1, σ_2 och Δ. Olikheter mellan $\mu_1, \ldots, \mu_n$ avspeglar olikheter mellan objekten, medan Δ anger den systematiska skillnaden mellan B:s och A:s mätningar. Om $\Delta > 0$ är B:s värden i genomsnitt större än A:s.

KTH VETENSKAP OCH KONST
KTH Matematik

Stickprov i par

Vi har alltså här antagit att det systematiska felet är oberoende av objektets verkliga värde, vilket kan vara realistiskt om det inte är för stora skillnader mellan objekten. Observera också att personerna inte behöver arbeta med samma precision (σ_1 och σ_2 tillåts vara olika).

KTH Matematik

Stickprov i par

Vanligen är man mest intresserad av Δ. För att få kunskap härom använder man ett trick, som bäst belyses av ett exempel.

KTH Matematik

Exempel på stickprov i par

På ett bryggeri gör man varje dag analyser av alkoholhalten i öl. Dessa analyser utförs av två kemister A och B. Man vill undersöka om det finns någon systematisk skillnad mellan As och Bs mätningar. Varje dag, under n dagar låter man A och B, oberoende av varandra, analysera samma prov.

KTH Matematik

Exempel på stickprov i par

Vi leds då till följande modell:

$$\begin{array}{llll} X_1, X_2, \ldots, X_n & \text{är} & N(\mu_i, \sigma_A) & \text{(As analyser)} \\ Y_1, Y_2, \ldots, Y_n & \text{är} & N(\mu_i + \Delta, \sigma_B) & \text{(Bs analyser)} \end{array}$$

där alla Xen och Yna är oberoende. Vi menar här att X_i är $N(\mu_i, \sigma_A)$-fördelad och att Y_i är $N(\mu_i + \Delta, \sigma_B)$-fördelad.

KTH Matematik

Exempel forts.

Tricket är nu att bilda

$$Z_i = Y_i - X_i$$

som är $N(\Delta, \sigma)$-fördelad, med

$$\sigma \left(= \sqrt{\sigma_A^2 + \sigma_B^2 - 2Cov} \right) .$$

KTH Matematik

Exempel forts.

Vi har nu återfört problemet till fallet med ett stickprov, och kan ge konfidensintervall för Δ på samma sätt som vi gjorde för μ, dvs.

$$\underline{I_\Delta = \overline{z} \pm \lambda_{\alpha/2}\sigma/\sqrt{n}}$$

om σ känd och

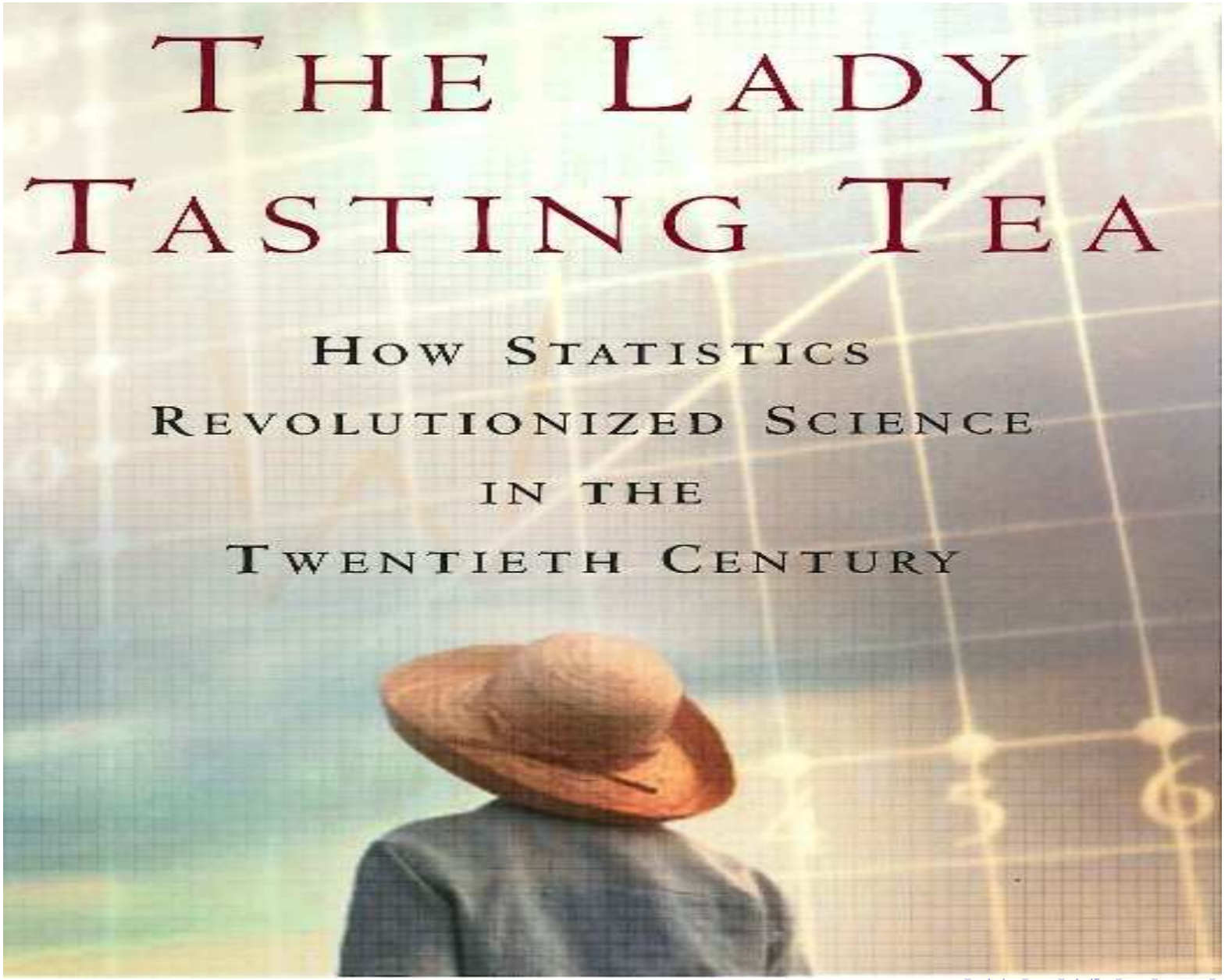
$$\underline{I_\Delta = \overline{z} \pm t_{\alpha/2}(n-1)s/\sqrt{n}}$$

om σ okänd, där

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \overline{z})^2.$$

KTH Matematik

Hypotesprövning



KTH Matematik

Hypotesprövning

- David Salsburg: The Lady Tasting Tea. How Statistics Revolutionized Science in the Twentieth Century. W.H. Freeman Co, New York 2001.

Denna populärvetenskapliga bok innehåller mycket mer än enbart hypotesprövning. I själva verket behandlas allt som ingår i denna kurs på ett underhållande vis.

KTH Matematik

Hypotesprövning

- Vi kommer att formulera en rätt så abstrakt svit av begrepp och modeller.
- Därför börjar vi med ett exempel.

KTH Matematik

Hypotesprövning: ESP

En person påstår att hen har extrasensory perception (ESP), som yttrar sig i förmåga att med förbundna ögon avgöra om krona eller klave kommer upp vid kast med ett mynt.

KTH Matematik

Hypotesprövning: ESP

- Låt p vara den okända sannolikheten att hen svarar rätt vid ett sådant kast.
- Man kastar ett symmetriskt mynt 12 gånger och med ledning av antalet korrekta svar x pröva vad som kallas nollhypotesen

$$H_0: \quad p = 1/2$$

(som innebär att personen bara gissar).

- Modellen: För varje p gäller att x är en observation av $X \in \text{Bin}(12, p)$ och speciellt, om H_0 är sann, att $X \in \text{Bin}(12, 1/2)$.

KTH VETENSKAP OCH KONST
KTH Matematik

Hypotesprövning: ESP

En procedur:

- Förkasta H_0, d.v.s påstå att personen har ESP, om x är tillräckligt stort, säg om $x \geq a$, men inte annars.
- Storheten a bör bestämmas så att sannolikheten är liten att H_0 förkastas om H_0 skulle vara sann. Därigenom garderar vi oss mot att påstå att personen har ESP om detta inte är sant.
- Angivna sannolikhet kallar vi **felrisken** eller **signifikansnivån** (ofta 0.05, 0.01 och 0.001)

KTH Matematik

Hypotesprövning: ESP

Börja med felrisken 0.05; vi säger vi att felrisken skall vara ca (men inte mer än) 0.05. Dvs. $P(X \geq a)$, om $p = 1/2$, bör vara ca 0.05, ty om $x \geq a$ i detta fall kommer vi felaktigt att påstå att H_0 är falsk, d.v.s att personen har ESP.

KTH Matematik

Hypotesprövning: ESP

$$P(X \geq a) = \sum_{i=a}^{12} \binom{12}{i} \left(\frac{1}{2}\right)^{12} \lesssim 0.05. \tag{1}$$

För att lösa denna ekvation i a kan man pröva sig fram.

12	11	10	9
0.00024	0.00293	0.01611	0.05371

För $a = 10$ blir summan 0.016 och närmare 0.05 kan man inte komma, eftersom man inte vill överskrida detta tal. Om personen svarar rätt minst 10 gånger bör man alltså påstå att han har ESP, men inte annars.

KTH Matematik

Hypotesprövning: ESP

Om man sänker felrisken från 0.05, först till 0.01, sedan till 0.001 :

$$\begin{array}{lll} \text{Felrisk} \lesssim & 0.05 & a = 10 \\ & 0.01 & 11 \\ & 0.001 & 12 \end{array}$$

I det sista fallet måste vi alltså kräva helt riktigt svar. Vi ser att det inte går att minska felrisken hur mycket som helst; skulle man vara så rädd för felaktigt uttalande att man vill ha en felrisk på, säg, 10^{-6}, måste man kasta myntet mer än 12 gånger.

KTH Matematik

Hypotesprövning: begrepp i samband med exemplet om ESP

- $X \in \text{Bin}(12, p)$: en testvariabel, x observerat värde på X
- $x \geq a$: ett kritiskt område (ett ensidigt test)
- $H_0 : \quad p = 1/2$ nollhypotes
- Beslutsregel: Förkasta H_0 om observationen hamnar i kritiskt område.
- Bestäm a så att $P(X \geq a) = \alpha$, $\alpha =$ felrisk, signifikansnivå.
- $P(X \geq a) = \sum_{i=a}^{12} \binom{12}{i} \left(\frac{1}{2}\right)^{12}$ om H_0 sann.

KTH Matematik

Hypotesprövning: alternativ hypotes

Nytt påstående: 'Jag kan i nio fall av tio svara rätt.' Detta är en mothypotes (hypotes mot H_0)

$$H_0: \quad p = 1/2$$

$$H_1: \quad p = 9/10$$

Tag $\alpha = 0.05$. Förkasta H_0 om $x \geq 10$.

- $h(0.9) = P(X \geq 10) = \sum_{i=10}^{12} \binom{12}{i} \left(\frac{9}{10}\right)^i \left(\frac{1}{10}\right)^{12-i}$
  Detta är sannolikheten för att H_0 förkastas om $p = 9/10$ sant. (=TESTETS STYRKA.)

$h(p) = \sum_{i=10}^{12} \binom{12}{i} \left(p\right)^i \left(1-p\right)^{12-i}$ kallas testets **styrkefunktion**.

Hypotesprövning: ESP

Du kan själv testa din egen förmåga för ESP med Rhines och Zeners test, som är ungefär som ovan men har fem symboler och räknar signifikans med t -fördelning (jfr. nedan). Se:

`http://www.scientificpsychic.com/esp/esptest.html`

KTH Matematik

Vi avslutar denna föreläsning genom att ta fram de generiska beståndsdelarna av varje tillfälle av hypotesprövning/statistiskt test samt tänket i statistiska test-

KTH Matematik

Rare Event Rule

If, under a given assumption, the probability of an observed event is very small, we conclude that the assumption is likely not correct.

KTH Matematik

The Basics of a Statistical Test

Definition

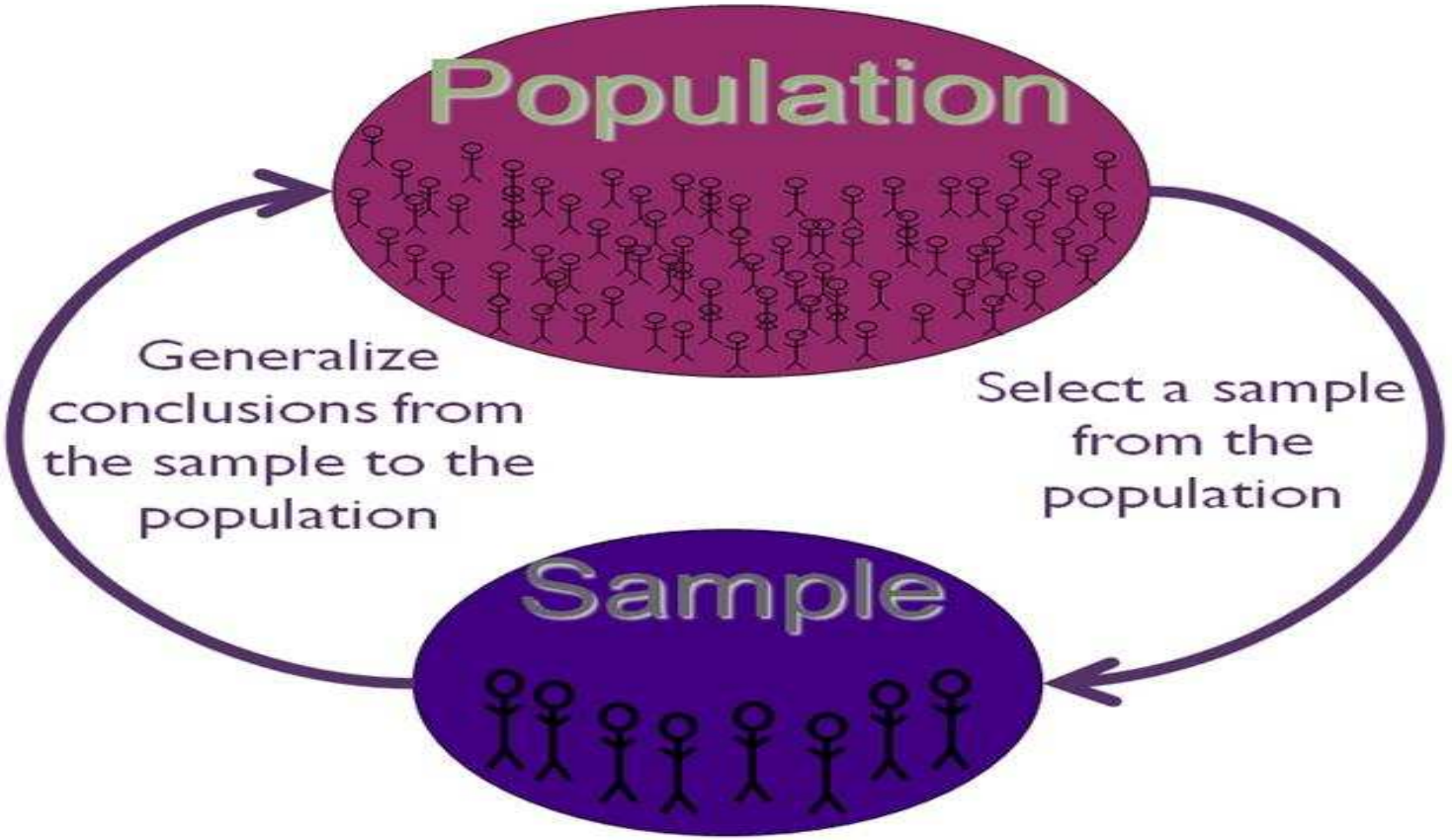
In statistics, a **hypothesis** *is a claim or statement about a property of a population. A* **hypothesis test** *(or a significance test) is a standard procedure for checking a claim about a property of a population.*

Example of a hypothesis:

- In genetics, G. Mendel claims that under certain circumstances, the percentage of the offspring peas with yellow pods is 25 %.

KTH VETENSKAP OCH KONST
KTH Matematik

A **hypothesis** is a claim or statement about a property of a population.



KTH Matematik

Learning Outcomes

Standard procedure contains: *null hypothesis, alternative hypothesis, test statistic, significance level, critical value, decision rule, proceed assuming null hypothesis is true, with the logic of the rule of rare event.*

KTH Matematik

The Individual Components of a Statistical Test

- Given a claim identify a statistical model for your population and the **null hypothesis** and the **alternative hypothesis**.
- Given a claim and sample data, compute the value of the **test statistic**
- Given **significance level**, identify the **critical values**.
- Given a value of a test statistic, identify significance level, identify the *p*-**value**.

We shall now explain what this, excluding for the moment *p*-value, is.

KTH Matematik

The Structure and Logic a Statistical Test: Null and Alternative Hypothesis

- The **null hypothesis** (denoted by H_o) is a statement that the value of a population parameter (such as proportion, mean, standard deviation) is equal to some claimed value. Examples:

$$H_0 : p = 0.5 \quad H_0 : \mu = 0.86$$

  We **assume that H_0 is true and reach a conclusion either to reject H_0 or fail to reject H_0.**

- The **alternative hypothesis** (denoted by H_1) is a statement that the value of a population parameter is somehow different from the null hypothesis.

$$H_1 : p > 0.5 \quad H_1 : p < 0.5 \quad H_1 : p \neq 0.5$$

KTH Matematik

The Structure and Logic a Statistical Test: Test Statistic

The **test statistic** is a value computed from the sample data and it is used to make the decision about the rejection of the null hypothesis. The test statistic is used to for determining whether there is significant evidence against the null hypothesis. Examples:

- Test statistic (=testvariabel) for proportion: $z = \frac{\widehat{p}-p}{\sqrt{\frac{pq}{n}}}$.
- Test statistic for mean: $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ or $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$.

KTH Matematik

The Structure and Logic of a Statistical Test: a Statistical Test: Critical Region, Significance Level

- The **critical region** (or **rejection region**) is the set of all values of the test statistic that cause us to reject the null hypothesis.
- The **significance level** denoted by α is the probability that the test statistic will fall in the critical region when the null hypothesis is actually true. Common choices of α are 0.05, 0.01 and 0.10.
- A **critical value** is any value that separates the critical region, where we reject the null hypothesis, from the values of the test statistic that do not lead to rejection of the null hypothesis.

KTH Matematik

Two-tailed, Left-tailed, Right-tailed

The *tails* in a density curve are the regions bounded by critical values. Some hypothesis tests are two-tailed, some are right-tailed and some are left-tailed.

- **Two-tailed test:** the critical region in two parts, significance level is a sum of two areas under the density curve. The significance level is divied equally between the two tails.
- **Left-tailed test:** the critical region in the left tail of the density curve.
- **Right-tailed test:** the critical region in the right tail of the density curve.

KTH Matematik

Hypotesprövning

- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes
    - Teststorhet (testvariabel)
    - Kritiskt område, Signifikansnivå,
    - Styrkefunktion
- Tvåsidiga test
- Hypotesprövning och konfidensintervall

KTH Matematik