Statistik för bioteknik SF1911
Föreläsning 11: Hypotesprövning och statistiska test del 2.
Timo Koski

TK

28.11.2017

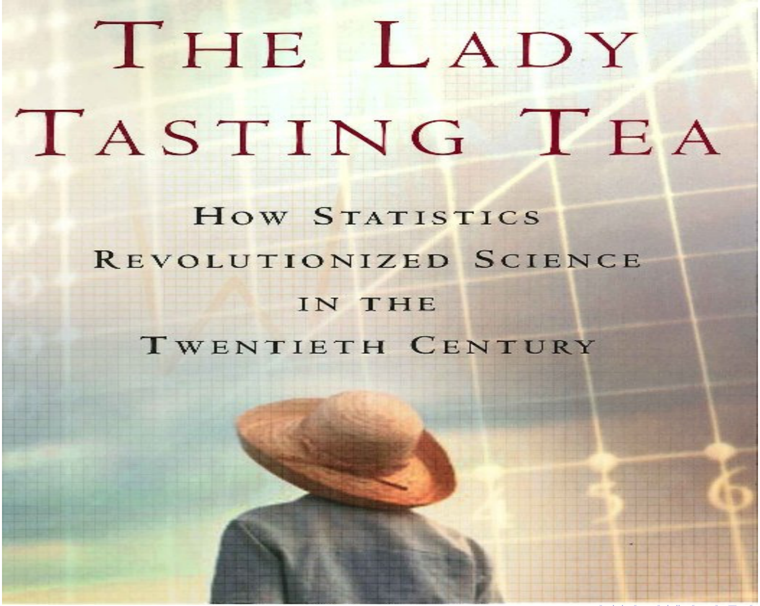
KTH Matematik

Outline of Lecture 11.

- Matched pairs or the paired samples (sticprov i par, parvisa observationer) t-test
- Type I error/Fel av typ I, Type II error/Fel av typ II
- Specificitet, sensitivitet
- Testets styrka/power
- Styrkefunktion/power function

KTH Matematik

Lady tasting tea



KTH Matematik

Lady tasting tea

David Salsburg: *Lady Tasting Tea - How Statistics Revolutionized Science in the Twentieth Century* Holt McDougal, 2002-05-01.
Brief commercial:
(The book is) saluting the spirit of those who dared to look at the world in a new way, this insightful, revealing history explores the magical mathematics that transformed the world.

KTH VETENSKAP OCH KONST
KTH Matematik

Matched Pairs

We shall deal with a testing situation " matched pairs " that is treacherously close to the comparison of two means 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

Matched Pairs/Stickprov i par

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.

Matched Pairs/Stickprov i par

- There is, as in case of two means, two series of observations. But the model for two means 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

Matched Pairs

We are primarily interested in Δ. To do the analysis we need a trick, which is best illustrated by another example.

Matched Pairs: Another example

A laboratory in a brewery takes daily samples of beer to analyse. Two chemists A and B analyse the alcoholic percentage in the samples. One asks if there was a systematic difference between A's and B's results. Daily, for n days, we let A and B, independently of each other, to analyse the same sample of beer, new sample per day.

Matched Pairs

Chemist	Beer sample			
	1	2	...	n
A	x_1	x_2	...	x_n
B	y_1	y_2	...	y_n

KTH Matematik

The Statistical Model:

$$
\begin{aligned}
X_1, X_2, \ldots, X_n &\sim N(\mu_i, \sigma_A) \quad \text{(A's results)} \\
Y_1, Y_2, \ldots, Y_n &\sim N(\mu_i + \Delta, \sigma_B) \quad \text{(B's results)}
\end{aligned}
$$

$$\Delta = \text{ a systematic difference}$$

KTH VETENSKAP OCH KONST
KTH Matematik

The trick

The trick is to form the **differences**

$$Z_i = Y_i - X_i$$

since then $Z_i \sim N(\Delta, \sigma)$ with $\sigma \left(= \sqrt{\sigma_A^2 + \sigma_B^2}\right)$. But now we have reduced the problem to a case with one sample and we can form confidence interval for Δ as we did for μ in Lect. 4.

KTH Matematik

Hypothesis testing for matched pairs

$\overline{z}$ is the mean value (=arithmetic mean of the samples z_i) of the differences $z_i = y_i - x_i$ of the matched pair data.

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

is the standard deviation for differences z_i of the matched data.

The test statistic is

$$t = \frac{\overline{z} - \Delta}{\frac{s}{\sqrt{n}}}$$

KTH Matematik

Hypothesis testing for matched pairs: critical region

The test statistic is

$$t = \frac{\bar{z} - \Delta}{\frac{s}{\sqrt{n}}}$$

If

$$H_o : \Delta = 0, \quad H_1 : \Delta \neq 0$$

then

$$t = \frac{\bar{z}}{\frac{s}{\sqrt{n}}}$$

The critical values $t_{\alpha/2}$ are obtained from a t-distribution with $n-1$ degrees of freedom.
The critical region is

$$t < -t_{\alpha/2} \quad \text{or} \quad t > t_{\alpha/2}$$

KTH VETENSKAP OCH KONST
KTH Matematik

Confidence Interval for Matched Pairs

$$\bar{z} - E < \Delta < \bar{z} + E$$

where

$$E = t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

If

$$H_o : \Delta = 0, \quad H_1 : \Delta \neq 0$$

Reject H_o if 0 is not in the interval.

KTH Matematik

Structure and Logic of Statistical Tests (General Discussion)

We take another look at the general principles.

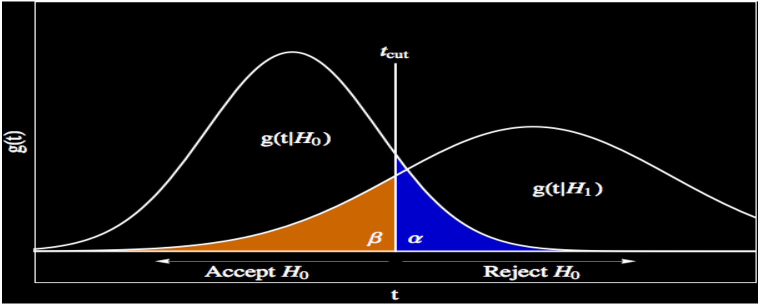
Structure and Logic of Statistical Tests (General Discussion)

When testing a null hypothesis, the statement is to either reject or fail to reject that null hypothesis. Such conclusions are sometimes correct and sometimes wrong, even if we do everything correctly in terms of computing margins of errors e.t.c., and have a good/perfect model of the population. *There are two types of errors:*

- **Type I error** *The mistake of rejecting the null hypothesis when it actually is true. The symbol α is used to represent the probability of type I error.*
- **Type II error** *The mistake of failing to reject the null hypothesis when it actually is false. The symbol β is used to represent the probability of type II error.*

KTH Matematik

Here we read $g(t \mid H_o)$ as the density curve in t, when H_0 is true, and $g(t \mid H_1)$ has a analogous meaning. Here, for e.g. $H_o : X \sim N(\mu, \sigma)$, $H_1 : X \sim N(\mu_1, \sigma_1)$, i.e., the alternative hypothesis is not a composite one. t_{cut} is the critical value.



Four possible outcomes of a test; c.f. rule of court

	Condition of null hypothesis	
Possible action	**True**	**False**
Fail to reject H_0	**Correct** $(1-\alpha)$	**Type II error** β
Reject H_0	**Type I error** α	**Correct** $(1-\beta)$

KTH Matematik

Controlling Type I error and Type II error (General Discussion)

- *For fixed* α *increase* n*, then* β *will decrease.*
- *For fixed* n*, decrease* α*, then* β *will increase.*
- *Increase* n*, decreases both* α *and* β*.*

KTH Matematik

Power of a Test

Definition

The **power** *of a hypothesis test is the probability* $1 - \beta$ *of rejecting a false null hypothesis. Power is computed by using a particular significance level* α, *a particular sample size* n, *the value of the population parameter used in the null hypothesis, and a particular value of the population parameter that is an alternative to the value assumed in the null hypothesis.*

KTH Matematik

Hypotesprövning: ESP

En person påstår att han 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.

Hypotesprövning: ESP

- Låt p vara den okända sannolikheten att han 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 Matematik

Hypoteseprövning: 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)

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.

Hypotesprövning: ESP

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

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.

Hypotesprövning: ESP

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

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

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.

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.)

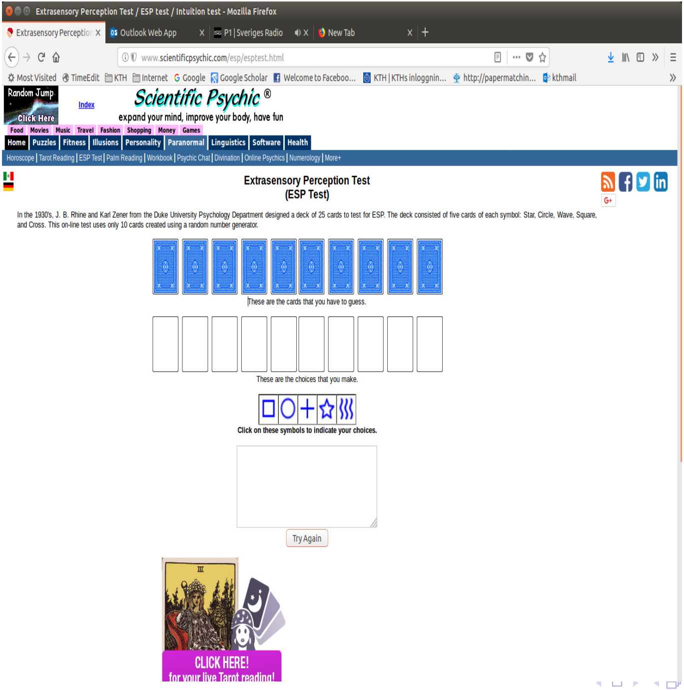
$h(p) = \sum_{i=10}^{12} \binom{12}{i} \left(p\right)^{i} \left(1-p\right)^{12-i}$ kallas testets **styrkefunktion**.

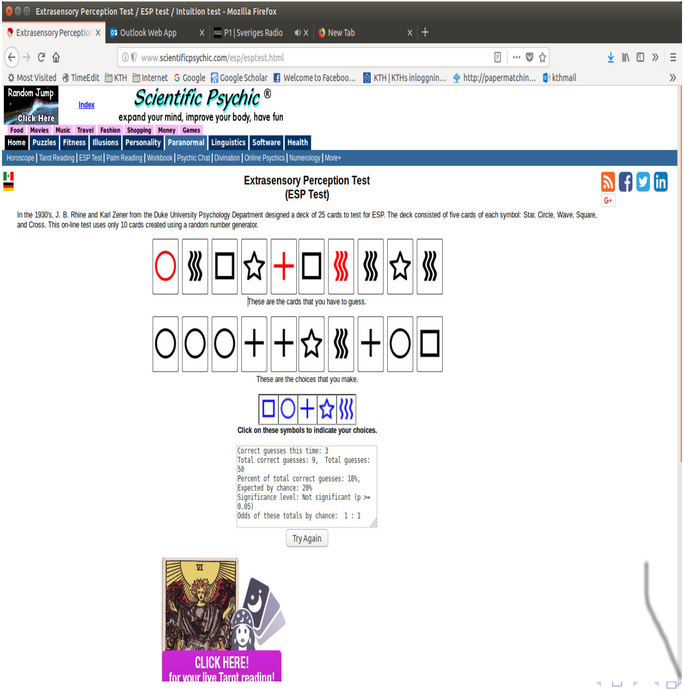
KTH Matematik

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`


Extrasensory Perception Test / ESP test / Intuition test - Mozilla Firefox
Extrasensory Perceptio
Outlook Web App
P1 | Sveriges Radio
New Tab
www.scientificpsychic.com/esp/esptest.html
Most Visited
TimeEdit
KTH
Internet
Google
Google Scholar
Welcome to Faceboo...
KTH | KTHs inloggnin...
http://papermatchin...
kthmail
Random Jump
Click Here
Index
Scientific Psychic ®
expand your mind, improve your body, have fun
Food Movies Music Travel Fashion Shopping Money Games
Home Puzzles Fitness Illusions Personality Paranormal Linguistics Software Health
Horoscope | Tarot Reading | ESP Test | Palm Reading | Workbook | Psychic Chat | Divination | Online Psychics | Numerology | More+
Extrasensory Perception Test
(ESP Test)
In the 1930's, J. B. Rhine and Karl Zener from the Duke University Psychology Department designed a deck of 25 cards to test for ESP. The deck consisted of five cards of each symbol: Star, Circle, Wave, Square, and Cross. This on-line test uses only 10 cards created using a random number generator.
These are the cards that you have to guess.
These are the choices that you make.
Click on these symbols to indicate your choices.
Try Again
III
CLICK HERE!
for your live Tarot reading!



KTH Matematik

$\widehat{Y}$ is an estimate of Y that has two values, 1 and $+1$. (E.g., in gene detection by a sequence model, $Y =$ gene or not.

False Positives = FP , True Positives = TP

False Negatives = FN , True Negatives= TN

	$Y = +1$	$Y = -1$
$\widehat{Y} = +1$	TP	FP
$\widehat{Y} = -1$	FN	TN

KTH Matematik

Error rates with statistics names

Let N=TN+FP, P=FN+TP.

- FP/N = type I error = 1−Specificity
- TP/P= 1-type II error =power =sensitivity=recall

Testyrka (power), Uppgift 9.2.9 i övningarna

Vi har n data som är $\mathcal{N}\left(\mu, \sigma^2\right)$. Vi testar

$$H_0 : \mu = 7$$

mot den alternativa hypotesen

$$H_A : \mu < 7$$

med hjälp av teststorheten (teststatistikan)

$$U = \frac{\bar{X} - 7}{\sigma / \sqrt{n}}.$$

Härvid förkastas H_0 om $u < -1.64$. Ensidigt test.

KTH Matematik

Testyrka (power), Uppgift 9.2.9 i övningarna

Testets styrka är sannolikheten att förkasta H_0 som funktion av parametern μ. Vi betecknar denna funktion med $h(\mu)$ och definierar i matematiska termer som

$$h(\mu) \stackrel{\text{def}}{=} P(U < -1.64; \mu) = P(\frac{\bar{X} - 7}{\sigma/\sqrt{n}} < -1.64; \mu)$$

Beteckningen $P(U < -1.64; \mu)$ innebär att vi beräknar denna sannolikhet med $\mathcal{N}\left(\mu, \sigma^2\right)$, där μ är något väntevärde $\mu < 7$.

$$= P(\frac{\bar{X}}{\sigma/\sqrt{n}} < \frac{7}{\sigma/\sqrt{n}} - 1.64)$$

$$= P(\underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{\sim\mathcal{N}(0,1)} < \frac{7 - \mu}{\sigma/\sqrt{n}} - 1.64)$$

KTH Matematik

$$= \Phi\left(\frac{7-\mu}{1.284/\sqrt{5}} - 1.64\right) =$$

$$= \Phi\left(\frac{7}{1.284/\sqrt{5}} - \frac{\mu}{1.284/\sqrt{5}} - 1.64\right)$$

$$= \Phi\left(12.1904 - 1.64 - 1.7415 \cdot \mu\right)$$

$$= \Phi\left(10.5504 - 1.7415 \cdot \mu\right)$$

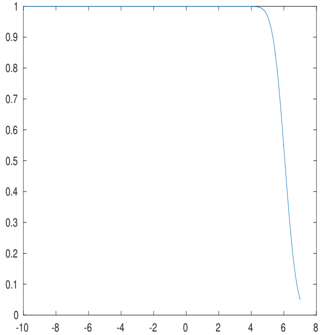
D.v.s.

$$h(\mu) = \Phi\left(10.5504 - 1.7415 \cdot \mu\right).$$

KTH Matematik

$$h(\mu) = \Phi\left(10.5504 - 1.7415 \cdot \mu\right).$$

Denna funktion återges grafiskt i figuren nedan:



Till exempel

$$h(4) = \Phi\left(10.5504 - 1.7415 \cdot \mu\right) = 0.9998$$

vilket säger att testets styrka $= 0.998$ för $\mu = 4$, eller, med andra ord att 0.9998 är sannolikheten att förkasta $H_o : \mu = 7$, när det sanna värdet på μ är 4.

$$h(7) = \Phi\left(10.5504 - 1.7415 \cdot 7\right) = 0.05.$$

är givetvis signifikansnivån, ty det kritiska värdet $-1.64 = -\lambda_{0.05}$.