

# TENTAMEN I TILLÄMPAD MATEMATIK OCH STATISTIK FÖR IT-FORENSIK. DEL 2: STATISTIK

## 7.5 HP

17 augusti, 2015 kl. 9.00 – 13.00

**Maxpoäng:** 30p. **Betygsgränser:** 12p: betyg 3, 18p: betyg 4, 24p: betyg 5.

**Hjälpmedel:** Miniräknare samt formelsamling som medföljer tentamenstexten.

**Kursansvarig:** Eric Järpe, telefon 0702-822 844, 035-16 76 53.

Till uppgifterna skall *fullständiga lösningar* lämnas. Lösningarna ska vara *utförligt* redovisade! Varje lösning ska börja överst på nytt papper. Endast en lösning per blad. Lösningar kommer finnas på internet: <http://dixon.hh.se/erja/teach> → Matematik och statistik för IT-forensik.

1. Beräkna sannolikheten  $P(2X + 3 \geq 1)$  då  $X \in N(-1, 5)$ . (3p)

2. Antag att man gör mätningarna

$x_i$	1	5	4	8	5
$y_i$	9	7	5	3	6

av variablerna  $X$  respektive  $Y$  på individerna 1, 2, 3, 4 och 5 (dvs ett parat stickprov). Beräkna

(a) stickprovsmedianen av  $X$ . (2p)

(b) förklaringsgraden för den linjära modellen med  $X$  som förklarande variabel och  $Y$  som responsvariabel. (4p)

3. Då Virve åker för att hälsa på släkt och bekanta händer det att hon kör fel, speciellt om det är många vägskäl på vägen hon färdas. Låt  $V$  vara antalet vägskäl på en godtycklig av dessa resor och låt  $F$  vara händelsen att Virve kör fel. Då är  $P(F|V=0) = 0$ ,  $P(F|V=1) = 0.02$ ,  $P(F|V=2) = 0.05$ ,  $P(F|V \geq 3) = 0.17$  och  $P(V=0) = 0.3$ ,  $P(V=1) = 0.55$ ,  $P(V=2) = 0.1$ ,  $P(V \geq 3) = 0.05$ . Beräkna

(a)  $P(F \cup \{V = 1\})$  (3p)

(b)  $P(V \geq 3 | F)$  (3p)

4. En textfil består av 1 830 tecken som delas in i 183 ej överlappande grupper med 10 tecken i varje grupp. I en slumpmässigt vald grupp är antalet förekomster av bokstaven "a" binomialfördelad med  $n = 10$  och  $p = 0.1$ . Vad är

(a) sannolikheten att en grupp innehåller minst 2 stycken "a"? (2p)

(b) (beräknat på en sida) den approximativa sannolikheten att det genomsnittliga antalet "a" per grupp är minst 1.1? (3p)

5. I en produktionsskog vill man avverka träden så snart den förväntade trädhöjden,  $X$ , överstiger 10 meter. Man mäter ett slumpmässigt urval av 100 träd och observerar

$$\sum_{i=1}^{100} x_i = 1\,051 \quad \sum_{i=1}^{100} x_i^2 = 11\,501$$

(a) Kan man på 5% signifikansnivå bevisa att det är dags för avverkning? (3p)

(b) Bilda ett 99% konfidensintervall för den förväntade trädhöjden. (3p)

6. En förgiftningsklinik får varje månad ta emot patienter. Under 24 månader observeras 88 fall fördelade enligt

<i>Antal patienter</i>	0	1	2	3	4	5	6	$\geq 7$
<i>Antal månader</i>	1	2	5	7	4	1	2	2

(dvs under 1 av de 24 månaderna är det inget förgiftningsfall, under [var och en av] 2 månader är det 1 förgiftningsfall, under [var och en av] 5 månader är det 2 förgiftningsfall, osv). Testa på 5% signifikansnivå om man kan förkasta påståendet att antalet förgiftningsfall per månad är Poissonfördelat med  $\lambda = 4$ . (4p)

*LYCKA TILL!*

# Matematik

## Definition 1 MÄNGDBETECKNINGAR

$\emptyset$  Tomma mängden     $\Omega$  Hela utfallsrummet  
 $\cup$  Unionen                 $\cap$  Snittet  
 $^C$  Komplementet         $|A|$  Antalet element i  $A$

## Sats 1 ADDITIONSSATSEN

För alla mängder  $A$  och  $B$  gäller att  $|A \cup B| = |A| + |B| - |A \cap B|$ .

## Sats 2 DE MORGANS LAGAR

För alla mängder  $A$  och  $B$  gäller att  $(A \cup B)^C = A^C \cap B^C$  och  $(A \cap B)^C = A^C \cup B^C$ .

## Sats 3 EXPONENTLAGARNA

$a^{b+c} = a^b a^c$ ,  $a^{bc} = (a^b)^c = (a^c)^b$ ,  $a^0 = 1$ ,  $a^1 = a$ ,  $a^{-1} = \frac{1}{a}$  och  $a^{1/2} = \sqrt{a}$ .

## Sats 4 LOGARITMLAGARNA För alla $a > 0$ , $b > 0$ , $c > 0$ och $d \in \mathbb{R}$ gäller

$\log_a(bc) = \log_a b + \log_a c$ ,  $\log_a(b^c) = c \log_a b$ ,  $\log_a a = 1$ ,  $\log_a 1 = 0$ ,  $\log_a \frac{b}{c} = \log_a b - \log_a c$ .

## Sats 5 KVADRERINGSREGLERNA

$(a+b)^2 = a^2 + 2ab + b^2$ ,  $(a-b)^2 = a^2 - 2ab + b^2$  och  $(a+b)(a-b) = a^2 - b^2$ .

## Sats 6 ANDRAGRADSEKVATIONER

Om  $x^2 + px + q = 0$  så är  $x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$ .

## Sats 7 FAKTORSATSEN

Varje polynom  $p(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} + a_nx^n$  av grad  $n$  har  $n$  nollställen  $x_1, x_2, \dots, x_n$  och kan faktoriseras mha dessa enligt  $p(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$ .

## Sats 8 SAMBANDET MELLAN KOEFFICIENTER OCH RATIONELLA RÖTTER

Om ekvationen

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n = 0$$

har en rationell rot  $x = p/q$  så måste  $a_0$  vara multipel av  $p$  och  $a_n$  vara multipel av  $q$ .

### Algoritm 1 DIVISIONSALGORITMEN

För alla heltal  $a$  och  $b \neq 0$  finns det heltal  $k$  och  $r$  sådana att  $0 \leq r \leq |b| - 1$  och

$$\frac{a}{b} = k + \frac{r}{b}$$

där talet  $k$  kallas **kvot** och talet  $r$  kallas **(principal) rest**.

### Definition 2

Ett **primtal** är ett heltal som inte är jämnt delbart med något annat heltal andra än 1 och sig självt.

### Algoritm 2 ERATOSTHENES SÅLL

Antag att man vill generera alla primtal  $\leq n$ .

1. Gör en lista över alla heltal från 2 till  $n$ .
2. Ringa in det första icke strukna eller inringade talet.
3. Stryk alla multipler av det senast inringade talet från resten av listan.
4. Om inte alla tal  $\leq \sqrt{n}$  är inringade eller strukna, gå tillbaka till steg 2.
5. Då alla tal som är  $\leq \sqrt{n}$  behandlats är de icke strukna talen primtalen.

### Definition 3

Den **största gemensamma delaren**,  $\gcd(a, b)$ , för två heltal,  $a$  och  $b$ , är produkten av alla primtalsfaktorer som är gemensamma i  $a$  och  $b$ .

### Definition 4

Heltalen  $a$  och  $b$  kallas **relativt prima** om  $\gcd(a, b) = 1$ .

### Algoritm 3 EUKLIDES ALGORITM

För att bestämma  $\gcd(a, b)$ , där  $a > b$ , bestäm  $r_1, r_2, r_3, \dots$  så att

$$\begin{cases} a = c_1b + r_1 & \text{där } 0 \leq r_1 \leq |b| - 1 \\ b = c_2r_1 + r_2 & \text{där } 0 \leq r_2 \leq r_1 - 1 \end{cases}$$

och fortsättningsvis

$$\begin{cases} r_1 = c_3r_2 + r_3 & \text{där } 0 \leq r_3 \leq r_2 - 1 \\ r_2 = c_4r_3 + r_4 & \text{där } 0 \leq r_4 \leq r_3 - 1 \\ \vdots & \vdots \\ r_{n-2} = c_nr_{n-1} + r_n & \text{där } 0 \leq r_n \leq r_{n-1} - 1 \\ r_{n-1} = c_nr_n + 0 & (\text{där alltså } r_{n+1} = 0) \end{cases}$$

Den första resten  $r_i$  som är  $= 0$  (dvs  $r_{n+1}$  i förklaringen ovan) kallas den **första försvinnande** resten, den senaste resten innan den ( $r_n$  i förklaringen ovan) kallas den **sista icke-försvinnande** resten. Och det är den sista icke-försvinnande resten som är  $\gcd(a, b)$ .

### Definition 5

Låt  $a$  och  $b$  vara heltal. Det minsta tal,  $c$ , sådant att  $a = bc$  eller  $b = ac$  kallas **minsta gemensamma multipel** för  $a$  och  $b$  och betecknas  $\text{lcm}(a, b)$ .

**Sats 9**  $\text{lcm}(a, b) = \frac{ab}{\text{gcd}(a, b)}$  för alla heltal  $a$  och  $b$ .

### Algoritm 4 LÖSNING AV DIOFANTISK EKVATION

För att lösa den diofantiska ekvationen  $ax + by = c$

1. beräkna  $d = \text{gcd}(a, b)$  mha Euklides algoritm.
2. Om inte  $c$  är en multipel av  $d$  så saknar ekvationen heltalslösningar.
3. Om  $c$  är en multipel av  $d$ , låt  $k = \frac{c}{d}$ .
4. Lös **hjälpkvationen**  $ax + by = d$  mha Euklides algoritm baklänges  $\Rightarrow (x_0, y_0)$ .
5. Allmän lösning till den fullständiga  $ax + by = c$  är då  $\{(kx_0 + bn, ky_0 - an), n \in \mathbb{Z}\}$ .

### Sats 10 RESTRÄKNING

Om  $a \equiv r$  och  $b \equiv s \pmod{c}$ , så är  $a + b \equiv r + s \pmod{c}$ .

Om  $a \equiv r$  och  $b \equiv s \pmod{c}$ , så är  $ab \equiv rs \pmod{c}$ .

Om  $a \equiv r \pmod{c}$ , så är  $a^b \equiv r^b \pmod{c}$ .

**Definition 6** Den diskreta (multiplikativa) inversen till  $x$  mod  $n$  är ett tal  $b$  som satisfierar  $ab \equiv 1 \pmod{n}$ .

**Definition 7** Den diskreta  $a$ -logaritmen till  $x$  mod  $n$  är ett tal  $b$  som satisfierar  $a^x \equiv b \pmod{n}$ .

### Algoritm 5 FERMATS FAKTORISERINGSMETOD

Antag att man vill faktorisera det udda talet  $N$ , dvs man vill hitta heltal,  $p$  och  $q$ , sådana att  $N = pq$ . Då kan man göra enligt följande procedur. Om talet man vill faktorisera är ett jämnt tal, bryt ut faktorn 2 och fortstt tills ett udda tal,  $N$ , erhålls.

1. Låt (initialt)  $x = 1 + \lceil \sqrt{N} \rceil$
2. Beräkna  $x^2 - N$ .
3. Om  $x^2 - N$  är en jämn kvadrat (dvs om  $\sqrt{x^2 - N}$  är ett heltal), låt  $p = x + \sqrt{x^2 - N}$  och  $q = x - \sqrt{x^2 - N}$  och gå till 6.
4. Om  $x - \sqrt{x^2 - N} < 2$ , låt  $p = N$  och  $q = 1$  och gå till 6.
5. Addera 1 till  $x$  och gå till 2.
6. Klart!

Om faktoriseringen blir  $p = N$  och  $q = 1$  (såsom det kan i steg 4. ovan) så är talet  $N$  ett primtal.

### Sats 11 SUMMERINGSREGLER

$$\sum_{k=1}^n a b_k = a \sum_{k=1}^n b_k \quad \sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k$$
$$\sum_{k=m}^n a = (n-m+1)a \quad \sum_{k=m}^n a_k = \sum_{k=1}^n a_k + \sum_{k=1}^{m-1} a_k$$

### Sats 12 SPECIELLA REGLER

$$\sum_{k=1}^n k = \frac{n(n+1)}{2} \quad \sum_{k=0}^n a^k = \frac{a^{n+1} - 1}{a - 1} \quad \text{om } a \neq 1 \quad \sum_{k=1}^n (a_k - a_{k-1}) = a_n - a_0$$

TABELL MED VÄRDEN PÅ  $\ln x$  FÖR  $x = 0.5, 0.55, 0.6, 0.65, \dots, 3.95$ .

$x$	+0.00	+0.05	+0.10	+0.15	+0.20	+0.25	+0.30	+0.35	+0.40	+0.45
0.5	-0.6931	-0.5978	-0.5108	-0.4308	-0.3567	-0.2877	-0.2231	-0.1625	-0.1054	-0.0513
1.0	0.0000	0.0488	0.0953	0.1398	0.1823	0.2231	0.2624	0.3001	0.3365	0.3716
1.5	0.4055	0.4383	0.4700	0.5008	0.5306	0.5596	0.5878	0.6152	0.6419	0.6678
2.0	0.6931	0.7178	0.7419	0.7655	0.7885	0.8109	0.8329	0.8544	0.8755	0.8961
2.5	0.9163	0.9361	0.9555	0.9746	0.9933	1.0116	1.0296	1.0473	1.0647	1.0818
3.0	1.0986	1.1151	1.1314	1.1474	1.1632	1.1787	1.1939	1.2090	1.2238	1.2384
3.5	1.2528	1.2669	1.2809	1.2947	1.3083	1.3218	1.3350	1.3481	1.3610	1.3737

### Definition 8

En funktion kallas **inversen** till funktionen  $f$  och betecknas  $f^{-1}$  om  $f^{-1}(f(x)) = x$  för alla  $x$  som  $f$  är definierad för.

### Sats 13 DERIVERINGSREGLER

Om  $f$  och  $g$  är funktioner av variabeln  $x$  och  $a$  en konstant så gäller

1.  $\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}$
2.  $\frac{d}{dx}(af) = a \frac{df}{dx}$
3.  $\frac{d}{dx}(a) = 0$
4.  $\frac{d}{dx}(x^n) = nx^{n-1}$  om  $n \neq 0$
5.  $\frac{d}{dx}(f \cdot g) = f \frac{dg}{dx} + g \frac{df}{dx}$
6.  $\frac{d}{dx}(e^f) = \frac{df}{dx} \cdot e^f$
7.  $\frac{d}{dx}(\ln x) = \frac{1}{x}$
8. Kedjeregeln:  $\frac{d}{dx}(f(g(x))) = \frac{df}{dx}(g(x)) \cdot \frac{dg}{dx}(x)$

**Sats 14** Om  $f$  är en deriverbar funktion så gäller att  $\frac{df}{dx}(x) < 0$  om och endast om  $f$  är avtagande genom  $x$ ,  $\frac{df}{dx}(x) > 0$  om och endast om  $f$  är växande genom  $x$ .

### Sats 15 BINOMIALKOEFFICIENTER

Antalet sätt att välja  $k$  element bland  $n$  möjliga (utan återläggning och utan hänsyn till ordningen) är

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{där} \quad n! = \prod_{j=1}^n j$$

### Sats 16 BINOMIALSATSSEN

För alla reella tal  $a$  och  $b$  och positiva heltal  $n$  är

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

## Matematisk statistik

### Definition 9 SANNOLIKHET

Om ett experiment har  $m$  möjliga utfall varav  $g$  är gynnsamma för händelsen  $A$ , så är sannolikheten för  $A$  vilket betecknas  $P(A) = g/m$ .

### Sats 17 KOMPLEMENTSAITSEN

$$P(A^C) = 1 - P(A)$$

### Sats 18 ADDITIONSSATSSEN

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

### Definition 10

En **slumpvariabel**,  $X$ , är en (vanligtvis numerisk) generalisering av ett experiment. Mha slumpvariabeln kan olika händelser formuleras som att  $X$  har vissa värden. En slumpvariabels **utfallsrum**,  $\Omega_X$ , är mängden av de värden som slumpvariabeln kan anta.

### Definition 11

$A$  och  $B$  är **oberoende** händelser om  $P(A \cap B) = P(A)P(B)$ .

Två slumpvariabler,  $X$  och  $Y$  med utfallsrum  $\Omega_X$  resp.  $\Omega_Y$ , är **oberoende** om  $P(X \in M_X, Y \in M_Y) = P(X \in M_X)P(Y \in M_Y)$  för alla  $M_X \subseteq \Omega_X$  och  $M_Y \subseteq \Omega_Y$ .

### Sats 19 BINOMIALFÖRDELNING

Om  $X = Y_1 + Y_2 + \dots + Y_n$  där  $P(Y_k = 1) = p$  och  $P(Y_k = 0) = 1 - p$  för alla  $k = 1, 2, \dots, n$  och variablerna  $Y_1, Y_2, \dots, Y_n$  är oberoende av varandra, så är  $\mathbf{X} \in \mathbf{Bin}(n, p)$  (dvs  $X$  är **binomialfördelad** med  $n$  och  $p$ ) vilket innebär att dess sannolikhetsfunktion är  $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$  där  $k \in \{0, 1, \dots, n\} = \Omega_X$ ,  $E(X) = np$  och  $V(X) = np(1-p)$ .

### Sats 20 POISSONFÖRDELNING

Om  $X$  är Poissonfördelad med intensitet  $\lambda$  betecknas detta  $X \in Poi(\lambda)$  och innebär att  $P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$  där  $x \in \{0, 1, 2, \dots\} = \Omega_X$ ,  $E(X) = \lambda$  och  $V(X) = \lambda$ . Dessutom gäller att  $X \in Poi(\lambda_X) \perp Y \in Poi(\lambda_Y) \Rightarrow X + Y \in Poi(\lambda_X + \lambda_Y)$ .

### Sats 21 NORMALFÖRDELNING

Denna betecknas  $N(\mu, \sigma^2)$  där  $\mu$  är väntevärde och  $\sigma^2$  är varians. Om  $X \in N(0, 1)$  kallas  $X$  **standard normalfördelad**, och dess fördelningsfunktion är  $\Phi(x) = P(X \leq x)$  för alla  $x \in \mathbb{R} = \Omega_X$ . Om  $X \in N(\mu, \sigma^2)$  så är  $P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$  för alla  $x \in \mathbb{R} = \Omega_X$ .

Symmetri:  $\Phi(-x) = 1 - \Phi(x)$  för alla  $x \in \mathbb{R}$ .

Sannolikheter:  $P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$  för all  $a < b \in \mathbb{R}$ .

**Definition 12 Väntevärdet** av en slumpvariabel  $X$  betecknas  $E(X)$  och är tyngdpunkten i sannolikhetsfunktionen respektive täthetsfunktionen för  $x$ . Linjaritet:  $E(aX + bY) = aE(X) + bE(Y)$ . **Variansen** av en slumpvariabel  $X$  betecknas  $V(X)$  och definieras  $V(X) = E((X - E(X))^2)$ . Räkneregel:  $V(X) = E(X^2) - E(X)^2$ . För diskreta variabler  $X$  är  $E(g(X)) = \sum_{x \in \Omega_X} g(x)P(X = x)$ .

### Sats 22 CENTRALA GRÄNSVÄRDESSATSEN (CGS)

Om  $X_1, X_2, \dots, X_n$  är oberoende och lika fördelade med  $E(X_i) = \mu$  och  $V(X_i) = \sigma^2$  så är approximativt  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \in N(\mu, \frac{\sigma^2}{n})$  och  $\sum_{i=1}^n X_i \in N(n\mu, n\sigma^2)$  då  $n$  är stort.

### Definition 13 BESKRIVANDE STATISTIK

**Proportion:**  $p = P(\widehat{X} \in A) = \frac{\#\{i : x_i \in A\}}{n}$

**Medelvärde:**  $\bar{x} = \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$

**Stickprovsvariens:**  $s^2 = \hat{\sigma}^2 \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$

**Stickprovskorrelation:**  $r^2 = \hat{\rho}^2 \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)}}$

### Definition 14 KONFIDENSINTERVALL

Antag  $X_1, X_2, \dots, X_n$  är stickprov på  $X$  och  $E(X) = \mu_X$ , att  $Y_1, Y_2, \dots, Y_m$  är stickprov på  $Y$  och  $E(Y) = \mu_Y$  och att  $V(X) = V(Y) = \sigma^2$ . Då gäller att ett  $100(1 - \alpha)\%$  konfidensintervall för

$\mu_X$  är  $\begin{cases} \bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} & \text{om } \sigma^2 \text{ är känd} \\ \bar{x} \pm t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} & \text{om } \sigma^2 \text{ är okänd} \end{cases}$

$\mu_X - \mu_Y$  är  $\begin{cases} \bar{x} - \bar{y} \pm t_{\alpha/2, (n+m-2)} s_P \\ \text{där } s_P^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right) \text{ om } \min(n, m) \leq 30 \\ \text{och } s_P^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m} \text{ om } \min(n, m) > 30 \end{cases}$



### Definition 15 HYPOTESTEST

Antag  $x_1, \dots, x_n$  är ett stickprov på  $X$  fördelad med parametern  $\theta$  respektive  $x_1, \dots, x_{n_1}$  och  $y_1, \dots, y_{n_2}$  på  $X$  och  $Y$  fördelade med parametern  $\theta$ . För att testa

$$\begin{cases} H_0 : \theta = \theta_0 & (\text{nollhypotesen}) \\ H_1 : \theta \in \Theta & (\text{alternativhypotesen}) \end{cases}$$

används teststatistikan  $U = U(X_1, \dots, X_n)$  och beslutsregeln  $A_\alpha$  som svarar mot  $\Theta$  enligt fördelningen av  $F_U$  under  $H_0$  vid signifikansnivån  $\alpha$ .

Testregeln är  $\begin{cases} \text{Förkasta } H_0 \text{ om } A_\alpha \\ \text{Förkasta inte } H_0 \text{ om inte } A_\alpha \end{cases}$

$\theta$	$H_0$	$H_1$	$u$	$A_\alpha$
$\pi$	$\pi = \pi_0$	$\pi < \pi_0$	$\frac{\sqrt{n}(p - \pi_0)}{\sqrt{\pi_0(1 - \pi_0)}}$ då $n\pi_0(1 - \pi_0) > 5$	$u < -\lambda_\alpha$
		$\pi > \pi_0$		$u > \lambda_\alpha$
		$\pi \neq \pi_0$		$ u  > \lambda_{\alpha/2}$
$\pi_1, \pi_2$	$\pi_1 = \pi_2$	$\pi_1 < \pi_2$	$\frac{p_1 - p_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$ då $n_1\pi_1(1 - \pi_1) > 5$ och $n_2\pi_2(1 - \pi_2)$	$u < -\lambda_\alpha$
		$\pi_1 > \pi_2$		$u > \lambda_\alpha$
		$\pi_1 \neq \pi_2$		$ u  > \lambda_{\alpha/2}$
$\mu$ ( $\sigma^2$ känd)	$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$u < -\lambda_\alpha$
		$\mu > \mu_0$		$u > \lambda_\alpha$
		$\mu \neq \mu_0$		$ u  > \lambda_{\alpha/2}$
$\mu$ ( $\sigma^2$ okänd)	$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$u < -t_\alpha(n - 1)$
		$\mu > \mu_0$		$u > t_\alpha(n - 1)$
		$\mu \neq \mu_0$		$ u  > t_{\alpha/2}(n - 1)$
$\mu_1, \mu_2$	$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$\frac{x_1 - x_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}(\frac{1}{n_1} + \frac{1}{n_2})}}$ då $\sigma_1 = \sigma_2$ men okända, och $\min(n_1, n_2) \leq 30$	$u < -t_{\alpha, (n_1+n_2-2)}$
		$\mu_1 > \mu_2$		$u > t_{\alpha, (n_1+n_2-2)}$
		$\mu_1 \neq \mu_2$		$ u  > t_{\alpha/2, (n_1+n_2-2)}$
$\mu_1, \mu_2$	$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$\frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ då $\sigma_1 = \sigma_2$ men okända, och $\min(n_1, n_2) > 30$	$u < -t_{\alpha, (n_1+n_2-2)}$
		$\mu_1 > \mu_2$		$u > t_{\alpha, (n_1+n_2-2)}$
		$\mu_1 \neq \mu_2$		$ u  > t_{\alpha/2, (n_1+n_2-2)}$
$A, B$	$A \perp B$	$A \not\perp B$	$\frac{(n_{11}c_2 - n_{12}c_1)\sqrt{n}}{\sqrt{c_1c_2r_1r_2}}$	$ u  > \lambda_{\alpha/2}$
$F_X$	$F_X = F_0$	$F_X \neq F_0$	$\sum_{k=1}^K \frac{(o_k - e_k)^2}{e_k}$ där $e_k = NP(X \in I_k)$	$u > \chi_\alpha^2(K - 1)$

### Enkel linjär regression

En linjär modell,  $Y = aX + b$ , som beskriver sambandet mellan slumpvariablerna  $X$  och  $Y$  baserad på det parade stickprovet

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  är med

$$\hat{a} = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad \text{och} \quad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

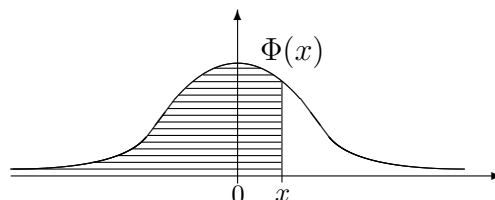
med förklaringsgraden

$$R^2 = \frac{\left( n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i) \right)^2}{\left( n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 \right) \left( n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2 \right)}$$

# Normalfördelningsvärden

Tabell över värden på  $\Phi(x) = P(X \leq x)$  där

$X \in N(0, 1)$ . För  $x < 0$  utnyttja relationen  $\Phi(x) = 1 - \Phi(-x)$ .



$x$	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

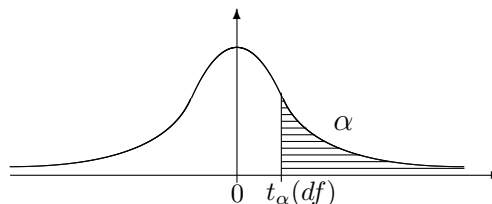
$x$	+0.0	+0.1	+0.2	+0.3	+0.4	+0.5	+0.6	+0.7	+0.8	+0.9
3	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

## Normal-percentiler:

Några värden på  $\lambda_\alpha$  sådana  
att  $P(X > \lambda_\alpha) = \alpha$   
där  $X \in N(0, 1)$

$\alpha$	$\lambda_\alpha$	$\alpha$	$\lambda_\alpha$
0.1	1.281552	0.005	2.575829
0.05	1.644854	0.001	3.090232
0.025	1.959964	0.0005	3.290527
0.01	2.326348	0.0001	3.719016

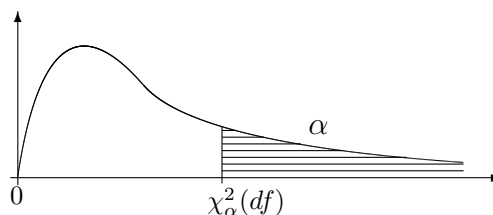
# $t$ -percentiler



Tabell över värden på  $t_\alpha(df)$ .

$df$	$\alpha$	0.25	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1		1.0000	3.0777	6.3138	12.7062	15.8945	31.8205	63.6567	318.3088
2		0.8165	1.8856	2.9200	4.3027	4.8487	6.9646	9.9248	22.3271
3		0.7649	1.6377	2.3534	3.1824	3.4819	4.5407	5.8409	10.2145
4		0.7407	1.5332	2.1318	2.7764	2.9986	3.7470	4.6041	7.1732
5		0.7267	1.4759	2.0150	2.5706	2.7565	3.3649	4.0322	5.8934
6		0.7176	1.4398	1.9432	2.4469	2.6122	3.1427	3.7074	5.2076
7		0.7111	1.4149	1.8946	2.3646	2.5168	2.9980	3.4995	4.7853
8		0.7064	1.3968	1.8595	2.3060	2.4490	2.8965	3.3554	4.5008
9		0.7027	1.3830	1.8331	2.2622	2.3984	2.8214	3.2498	4.2968
10		0.6998	1.3722	1.8125	2.2281	2.3593	2.7638	3.1693	4.1437
12		0.6955	1.3562	1.7823	2.1788	2.3027	2.6810	3.0545	3.9296
14		0.6924	1.3450	1.7613	2.1448	2.2638	2.6245	2.9768	3.7874
17		0.6892	1.3334	1.7396	2.1098	2.2238	2.5669	2.8982	3.6458
20		0.6870	1.3253	1.7247	2.0860	2.1967	2.5280	2.8453	3.5518
25		0.6844	1.3163	1.7081	2.0595	2.1666	2.4851	2.7874	3.4502
30		0.6828	1.3104	1.6973	2.0423	2.1470	2.4573	2.7500	3.3852
50		0.6794	1.2987	1.6759	2.0086	2.1087	2.4033	2.6778	3.2614
100		0.6770	1.2901	1.6602	1.9840	2.0809	2.3642	2.6259	3.1737

# $\chi^2$ -percentiler



Tabell över värden på  $\chi_\alpha^2(df)$ .

$df$	$\alpha$	0.999	0.995	0.99	0.95	0.05	0.01	0.005	0.001
1		0.0000	0.0000	0.0002	0.0039	3.8415	6.6349	7.8794	10.8276
2		0.0020	0.0100	0.0201	0.1026	5.9915	9.2103	10.5966	13.8155
3		0.0243	0.0717	0.1148	0.3518	7.8147	11.3449	12.8382	16.2662
4		0.0908	0.2070	0.2971	0.7107	9.4877	13.2767	14.8603	18.4668
5		0.2102	0.4117	0.5543	1.1455	11.0705	15.0863	16.7496	20.5150
6		0.3811	0.6757	0.8721	1.6354	12.5916	16.8119	18.5476	22.4577
7		0.5985	0.9893	1.2390	2.1673	14.0671	18.4753	20.2777	24.3219
8		0.8571	1.3444	1.6465	2.7326	15.5073	20.0902	21.9550	26.1245
9		1.1519	1.7349	2.0879	3.3251	16.9190	21.6660	23.5894	27.8772
10		1.4787	2.1559	2.5582	3.9403	18.3070	23.2093	25.1882	29.5883
12		2.2142	3.0738	3.5706	5.2260	21.0261	26.2170	28.2995	32.9095
14		3.0407	4.0747	4.6604	6.5706	23.6848	29.1412	31.3193	36.1233
17		4.4161	5.6972	6.4078	8.6718	27.5871	33.4087	35.7185	40.7902
20		5.9210	7.4338	8.2604	10.8508	31.4104	37.5662	39.9968	45.3147
25		8.6493	10.5197	11.5240	14.6114	37.6525	44.3141	46.9279	52.6197
30		11.5880	13.7867	14.9535	18.4927	43.7730	50.8922	53.6720	59.7031
50		24.6739	27.9907	29.7067	34.7643	67.5048	76.1539	79.4900	86.6608
100		61.9179	67.3276	70.0649	77.9295	124.342	135.807	140.169	149.449