

**TENTAMEN:** Matematisk statistik för K (TMA072)

Onsdagen den 28 maj 2008, kl 14:00-18:00

**Lärare och jour:** Aila Särkkä, telefon 772 3542

**Hjälpmedel:** Formelsamling, tabeller (även BETA, Physics Handbook, skoltabeller, t.ex. TEFYMA), valfri miniräknare.

- 1) Det finns 500 studenter, som läser matte och fysik på ett universitet. Av dessa studenter fick 82 en femma i matte, 73 femma i fysik och 42 femma både i matte och i fysik.
  - a) Hitta sannolikheten att en student vald på måfå fick en femma minst i ett av de två ämnena.
  - b) Hitta sannolikheten att en student vald på måfå fick en femma i matte men inte i fysik.
  - c) Hitta sannolikheten att en student vald på måfå fick mindre än en femma minst i ett av de två ämnena. (3p)
  
- 2) Man undersöker effekter som uppstår när barn utsätts för kokain före födelsen och därför testades barnens matematiska förmåga när de var 4 år gamla. I studien ingår 190 barn till kokainmissbrukare och 186 barn, som inte hade utsatts till kokain. Från datan, som man har samlat in, har man beräknat att barnen som hade utsatts för kokain hade ett medelpoängtal 7.3 och standardavvikelse 3.0, och barnen som inte hade utsatts för kokain hade medelpoängtal 8.2 och standardavvikelse 3.0.
  - a) Använd signifikansnivån 0.01 för att testa om utsättande för kokain före födelsen är associerad med lägre poängtal för fyraåringar. Vilka antaganden har du gjort?
  - b) Konstruera ett (dubbelsidigt) 99% konfidensintervall för skillnaden mellan det genomsnittliga poängtal för barnen till kokainmissbrukare och det genomsnittliga poängtal för barnen utan kokainmissbruk. Vilka antaganden har du gjort? Verkar det vara skillnad mellan de två grupperna?
  - c) Vilka tre faktorer påverkar generellt längden av konfidensintervallen och hur?
  - d) Borde a) och b) ge samma resultat? Motivera. (7p)

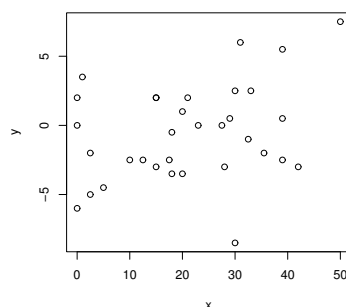
- 3) Man undersöker ett nytt läkemedels effekt på 50 olika faktorer genom att beräkna 50 stycken 95% konfidensintervall för dessa faktorerers genomsnittliga värden (baserad på stickprovsstorlek 40 vardera). Antag att läkemedlet inte har effekt på någon av de 50 faktorerna (dvs. att samtliga intervall borde innehålla det "normala" värdet som faktorn har utan något läkemedel). Hur stor är sannolikheten att trots allt minst två av dessa 50 konfidensintervall inte innehåller det "normala" värdet? (4p)
- 4) Man har mätt den lägsta temperaturen ( $X$ ) för 12 olika dagar och jämfört den med den lägsta temperaturen, som väderprognosen har givit fem dagar innan ( $Y$ ) (se tabellen nedan).

												Mean	SD	
$x$ :	14	14	15	20	24	21	19	22	23	21	19	18	19.17	3.38
$y$ :	16	17	19	16	24	22	21	24	25	22	21	21	20.67	3.08

- a) Utför ett hypotestest för att testa om det finns skillnad mellan de faktiska och förutsagda (väderprognosen) temperaturerna. Använd signifikansnivå 0.05 och ange antagandena som du har gjort för att kunna använda testet.
- b) Säg att den sanna skillnaden av de genomsnittliga värdena av  $X$  och  $Y$  var -1, dvs. att  $\mu_X - \mu_Y = -1$ . Vad är styrkan (power) av testet i a) då? (Det räcker att du ger ett intervall där teststyrkan ligger.)(5p)
- 5) a) Säg att du vill utföra ett hypotestest angående väntevärdet  $\mu$  och din nollhypotes är  $H_0 : \mu = \mu_0$ . Hur bestämmer du om du skall testa  $H_0$  mot  $H_1 : \mu \neq \mu_0$ ,  $H_1 : \mu < \mu_0$  eller  $H_1 : \mu > \mu_0$ ? Kan man använda datan man har samlat in för att bestämma detta?
- b) I hypotestest vill man att sannolikheten för typ I fel är liten och att sannolikheten för typ II fel är liten. Hur kan man garantera att både två är någonlunda små? (3p)

6) Botaniker vill undersöka matvanor för snögäss. Forskarna gav ingen mat till gässen tills deras magsäckar var tomma och lät dem sedan äta i sex timmar. Varje gås vägdes med tom mage och 2.5 timmar efter den hade börjat äta. Viktändringen rapporterades som ett procenttal av den ursprungliga vikten av gåsen. Man mätte också smältningseffektiviteten (i procenttal). Man ville undersöka sambandet mellan viktändringen och smältningseffektiviteten.

- a) Botanikerna var intresserade av korrelationen mellan viktändringen ( $Y$ ) och smältningseffektiviteten ( $X$ ). Tolka scatterplotten nedan, som är baserad på 33 gäss. Hurdan korrelationskoefficient (positiv/negativ, stort/litet värde) skulle du förvänta dig att få? Varför?

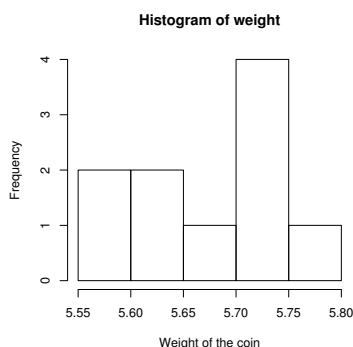


- b) Beräkna Pearsons korrelationskoefficient mellan viktändringen och smältningseffektiviteten ( $\sum_{i=1}^{33} x_i = 701.5$ ,  $\sum_{i=1}^{33} y_i = -18$ ,  $\sum_{i=1}^{33} x_i^2 = 21068.75$ ,  $\sum_{i=1}^{33} y_i^2 = 404$ ,  $\sum_{i=1}^{33} x_i y_i = 99.5$ ). Tolka värdet.
- c) När är det lämpligt att beräkna korrelationskoefficienten och när är det lämpligt att använda linjär regression? (3p)

- 7) Nedan har man viktena (i gram) och histogrammet av viktena ( $x_i$ ) av 10 likadana mynt valda på måfå:

5.7027 5.7495 5.7050 5.5941 5.7247 5.6114 5.6160 5.5999 5.7790 5.6841

Från data ovan kan man beräkna att medelvikten är 5.6766 och standardavvikelsen 0.0669.



- a) Vilket test skulle man kunna använda (givet att antagandena gäller) för att testa om medelvikten är 5.67g? Genomför testet på signifikansnivån 0.1 och lista antaganden som måste gälla för att kunna utföra det.
- b) Vilket test skulle man kunna använda (givet att antagandena gäller) för att testa om medianvikten är 5.67g? Genomför testet på signifikansnivån 0.1 och lista antaganden som måste gälla för att kunna utföra det.
- c) Skulle du rekommendera testet i a) eller i b)? Varför?
- d) Ibland måste man behandla de observerade värdena som är samma som  $H_0$ -värdet på ett speciellt sätt. För vilket test (i a) eller i b)) om för något gäller detta och varför? Hur kan man behandla de värdena (som är samma som  $H_0$ -värdet) då? (5p)

**Lycka till!**