Väl motiverade lösningar med svar krävs på varje uppgift, t.ex. ska modeller, approximationer, hypoteser och slutsatser anges och motiveras.

Institutionens papper används både som kladdpapper och som inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas. Skriv fullständigt namn på alla papper.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, Formelsamling "MASB11 Biostatistisk grundkurs" samt miniräknare.

Totalt kan man få 100 poäng. Gränsen för godkänd är 50 poäng och för väl godkänd 75 poäng.

Resultatet anslås senast 2011-08-31 på kursens hemsida www.maths.lth.se/matstat/kurser/masb11/vtm3/ samt i matematikhusets entréhall.

---

1. (a) Vikten hos en slumpmässigt vald 10-årig flicka i Sverige anses vara normalfördelad med väntevärde 34.5 kg och standardavvikelse 3.75 kg. Hur stor andel av de 10-åriga flickorna väger över 42 kg? (4p)

   (b) Fortsättning från 1(a). Tjugofem flickor väljs ut och vi beräknar deras medelvikt. Vad är sannolikheten att medelvärdet understiger 33 kg? (4p)

   (c) I en population är 40% kvinnor. Av kvinnorna är 10% vegetarianer medan samma siffra för männen är 8%. Vi väljer slumpmässigt ut en person ur populationen. Vad är sannolikheten att vi valt en vegetarian? (4p)

   (d) Låt X vara antalet nya AIDS-fall per dag vid ett stort sjukhus. Antag att sannolikhetsfunktionen för X ges av

x	0	1	2	3	4	5	6
$f(x)$	0.1	0.1	0.1	0.3	0.2	0.1	0.1

   Beräkna sannolikheten att man på en slumpmässigt vald dag får

   i. högst tre nya AIDS-fall
   ii. minst ett nytt AIDS-fall. (4p)

   (e) Antalet självmord en februarimånad i ett område är poissonfördelat med väntevärde $\lambda = 2$. Vad är sannolikheten att inga självmord inträffar de två närmaste februarimånaderna i detta område? (4p)

2. Vid en klinisk prövning av ett nytt läkemedel (B) som var tänkt att lindra andningssvårigheterna hos patienter med Kronisk Obstruktiv Lungsjukdom (KOL) undersökte man 143 slumpvis utvalda patienter. Av dessa behandlades 72 stycken med det etablerade preparatet A och resterande 71 patienter med det nya läkemedlet B. I den behandlingsgrupp som fick preparat A misslyckades behandlingen för 28 patienter och i den andra gruppen misslyckades behandlingen för 19 patienter.

Resultat	Misslyckad behandling	Lyckad behandling
Preparat A	28	44
Preparat B	19	52

   Avgör på lämpligt sätt om det finns någon signifikant skillnad i behandlingsresultat mellan preparat A och preparat B. (20p)

3. I en undersökning av metaller i biota mätte man ett år halten Cd (mg/kg) i lever och njure på 12 älgar i Kronobergs län. Samtidigt bedömdes älgens ålder. Resultat för Cd-halt i lever:

Älg nr	1	2	3	4	5	6	7	8	9	10	11	12
Ålder (år)	0.5	1.5	0.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	0.5	1.5
Cd-halt (mg/kg)	0.15	0.69	0.37	0.51	0.30	0.22	0.56	0.96	1.10	0.25	0.19	0.33

   Hur stor är skillnaden i förväntad Cd-halt i lever mellan älgar av ålder 0.5 år och 1.5 år? Svara i form av ett lämpligt konfidensintervall. Du får anta lämplig(a) normalfördelningar. (20p)

4. Ett sätt att mäta ett reningsverks kapacitet är att mäta värdet på Biochemical Oxygen Demand (BOD) i det renade vattnet. BOD-värdet mäts i mg syrgas per liter och ju lägre värde på BOD, desto effektivare är reningen. Tillåtet utsläpp från reningsverk ligger vanligen på maximalt 10 mg/l.

   (a) I ett reningsverk mättes BOD-värdet på avfallsvattnet vid sju olika tillfällen:

Tillfälle	1	2	3	4	5	6	7
BOD (mg/l)	7.1	9.5	5.3	8.6	10.7	12.7	10.3

   Ligger det förväntade BOD-värdet under gränsvärdet 10 mg/l? Antag lämplig(a) normalfördelningar. (10p)

(b) För att förbättra reningsverkets kapacitet prövar man en ny typ av rening. Vid ett test låter man en del av dagens avfallsvatten renas med den gamla metoden medan resten renas med den nya, varefter man mäter BOD-värdet. Dessa tester utfördes ungefär en gång varannan vecka i några månaders tid och gjordes då på den aktuella dagens avfallsvatten. Tyvärr gick det inte att få något BOD-värde för den nya reningstekniken den 24/2.

Testdag	3/2	17/2	24/2	3/3	14/3	21/3
BOD (mg/l) med gammal metod	5.6	12.4	7.3	10.3	2.6	9.5
BOD (mg/l) med ny metod	4.7	9.3	–	8.5	1.1	8.2

Ger den nya tekniken en signifikant förbättring av reningen så att förväntad BOD-värde blir lägre? Antag lämplig(a) normalfördelningar. (10p)

5. Bensen sprids i luften från bland annat bensindrivna fordon. Efter att ha infört katalytisk avgasrening och striktare regler för hur mycket bensen som får användas i bensinen har bensenhalten i luft, μg/m^3, i tätortsmiljö sjunkit sedan början av 1990-talet. I nedanstående tabell från Naturvårdsverket finns uppmätt bensenhalt i en svensk medeltätort för perioden 1994–2007, ($t_1 = 1$ motsvarar år 1994).

t_i	1	2	3	4	5	6	7
Bensen y_i (μg/m^3)	3.47	2.79	2.93	2.28	2.08	2.44	1.92

t_i	8	9	10	11	12	13	14
Bensen y_i (μg/m^3)	1.95	2.02	1.94	1.41	1.22	1.33	0.99

Räknehjälp:

$\sum t_i = 105$; $\sum y_i = 28.77$;
$SS_t = \sum(t_i - \bar{t})^2 = 227.5$;
$SP_{ty} = \sum(t_i - \bar{t})(y_i - \bar{y}) = -35.855$;
$SS_y = \sum(y_i - \bar{y})^2 = 6.3244$.

(a) Ansätt en linjär regressionsmodell $y_i = \beta_0 + \beta_1 \cdot t_i + \varepsilon_i$, $\varepsilon_i \in N\left(0,\ \sigma^2\right)$, för hur bensenhalten varierar med tiden och skatta alla tre parametrarna i modellen. (7p)

(b) Avgör på lämpligt sätt om förändringen av bensenhalten med avseende på tiden är signifikant. (6p)

(c) Utifrån dessa data, vad är den förväntade bensenhalten för år 2008? Svara med ett lämpligt intervall. (7p)

Lycka till!

Matematisk statistik
Lunds universitet

Kortfattade lösningar till tentamen i
MASB11-Biostatistisk grundkurs,
2011-08-25

1. (a) X=vikten $\sim N(34.5,\ 3.75^2)$. $P(X > 42) = 1 - P(X \leq 42) = 1 - P(Z \leq \frac{42-34.5}{3.75}) = 1 - P(Z \leq 2.0) = 1 - 0.9772 = 0.0228$. Det är alltså ca 2.3 % av flickorna som har en vikt över 42 kg.

   (b) $\overline{X}$=medelvärdet av 25 vikter. $\overline{X} \sim N(34.5,\ \frac{3.75^2}{25})$. $P(\overline{X} < 33) = P(Z < \frac{33-34.5}{\frac{3.75}{5}}) = P(Z < -2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$.

   (c) P(veg)=P(kvinna)·P(veg|kvinna)+P(man)·P(veg|man)=$0.4 \cdot 0.1 + 0.6 \cdot 0.08 = 0.088$.

   (d) X=antalet nya fall. $P(X \leq 3) = 0.1 + 0.1 + 0.1 + 0.3 = 0.6$. $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.1 = 0.9$.

   (e) X="antal självmord"; $X \sim Po(2)$. P(inget självmord en februarimånad)=$P(X = 0) = e^{-2}\frac{2^0}{0!} = e^{-2} = 0.1353$. (Här har vi använt att 0!=1.) Vi söker P(inget självmord under två oberoende månader)=$P(X = 0)^2 = 0.1353^2 = 0.0183$.

2. Detta kan behandlas som kategoridata och intressanta hypoteser är H_0 : "inget samband mellan preparat och behandlingsresultat"; H_1: "samband finns". Om H_0 är sann hade vi förväntat oss följande värden

	Misslyckad behandling	Lyckad behandling	Totalt
Preparat A	$\frac{47}{143} \cdot 72 = 23.66$	$\frac{96}{143} \cdot 72 = 48.34$	72
Preparat B	$\frac{47}{143} \cdot 71 = 23.34$	$\frac{96}{143} \cdot 71 = 47.66$	71
Totalt	47	96	143

$\chi^2 = \frac{(28-23.66)^2}{23.66} + \frac{(44-48.34)^2}{48.34} + \frac{(19-23.34)^2}{23.34} + \frac{(52-47.66)^2}{47.66} = 2.38.$

Eftersom $\chi^2 = 2.4 < 3.841 = \chi^2_{0.95,1}$ kan H_0 ej förkastas på nivå 0.05. Vi har inte påvisat att det finns något samband mellan preparat och behandlingsresultat.

3. Modell:

   Cd-halt från älgar med ålder 0.5 år, $x_1, \ldots, x_7$ observationer av $X \sim N(\mu_1,\ \sigma^2)$

   Cd-halt från älgar med ålder 1.5 år, $y_1, \ldots, y_5$ observationer av $Y \sim N(\mu_2,\ \sigma^2)$

   Med hjälp av räknare fås

   $\bar{x} = 0.2843,\ \ s_x = 0.123$

   $\bar{y} = 0.7280,\ \ s_y = 0.3083$

   Eftersom de två σ-skattningarna är någorlunda lika anser vi att antagandet om lika σ^2 i modellen är OK och vi gör en gemensam skattning av σ^2:

   $\hat{\sigma} = s_p = \sqrt{\frac{(7-1)s_x^2+(5-1)s_y^2}{(7-1)+(5-1)}} = 0.2170$

   Lämpliga hypoteser:

$H_0 : \mu_1 = \mu_2$, dvs ingen skillnad i genomsnittlig Cd-halt

$H_0 : \mu_1 \neq \mu_2$, dvs skillnad i genomsnittlig Cd-halt

Gör ett konfidensintervall för $\mu_1 - \mu_2$:

$I_{\mu_1 - \mu_2} = (\bar{x} - \bar{y} \pm t_{0.975,10} s_p \sqrt{\frac{1}{7} + \frac{1}{5}}) = (0.2843 - 0.7280 \pm 2.23 \cdot 0.2170 \sqrt{\frac{1}{7} + \frac{1}{5}}) =$ $(-0.7269,\ -0.1606)$.

Eftersom intervallet ej täcker över 0, kan H_0 förkastas på nivå 0.05. Det finns en skillnad mellan förväntad Cd-halt mellan åldersgrupperna. Älgar med ålder 1.5 år har i genomsnitt mellan $(0.16, 0.73)$ enheter högre Cd-halt i lever än älgar av ålder 0.5 år.

4. (a) Modell: $x_1, \ldots, x_7$ observationer av ξ=BOD-värde; $X \sim N(\mu,\ \sigma^2)$. Från data: $\hat{\mu} = \bar{x} = 9.1714$, $\hat{\sigma} = s = 2.4418$.

   Intressanta hypoteser är $H_0 : \mu = 10$; $H_1 : \mu < 10$ (gränsvärde understiget). Ett övre begränsat intervall för μ ges av $I_\mu = (-\infty,\ \bar{x} + t_{0.95,6} \cdot \frac{s}{\sqrt{7}}) = (-\infty,\ 10.965)$.

   Eftersom intervallet täcker över 10 kan H_0 inte förkastas på nivå 0.05. Från dessa data inte visat att förväntat BOD-värde signifikant ligger under gränsvärdet 10 mg/l.

   (b) Detta är en situation som bör behandlas med "stickprov i par". Eftersom vi inte har fullständiga data från tidpunkten 24/2 stryks mätningen med "BOD-värdet med gammal metod". För de återstående fem paren har vi följande modell för differenserna z_i="BOD-värde gammal metod vid tidpunkt i"-"BOD-värde ny metod vid tidpunkt i": 0.9, 3.1, 1.8, 1.5, 1.3 är observationer av ξ="differens i BOD-värde"; $\xi \sim N(\Delta,\ \sigma)$. Från data får vi att $\hat{\Delta} = \bar{z} = 1.72$, $\hat{\sigma} = s_z = 0.8379$.

   Intressanta hypoteser är $H_0 : \Delta = 0$ (ingen skillnad mellan metoder); $H_1 : \Delta > 0$ (ny metod ger i genomsnitt lägre BOD-värde).

   Ett undre begränsat intervall för Δ ges av $I_\Delta = (\bar{z} - t_{0.95,4} \cdot \frac{s_z}{\sqrt{5}},\ \infty) = (0.92,\ \infty)$.

   Eftersom intervallet inte täcker över 0 förkastas H_0 på nivå 0.05. Ja, det tycks som om den nya tekniken förbättrar reningen.

5. (a) $\hat{\beta}_1 = \frac{SP_{ty}}{SS_t} = \frac{-35.855}{227.5} = -0.1576$.
   $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{t} = 3.237$
   $\hat{\sigma} = \sqrt{\frac{1}{14-2}(SS_y - \frac{SP_{ty}^2}{SS_t})} = \sqrt{\frac{0.6734}{14-2}} = 0.2369$.

   (b) Ett 95% intervall för β_1 fås genom $I_{\beta_1} = (\hat{\beta}_1 \pm t_{0.975,12} \frac{\hat{\sigma^2}}{SS_t}) = (-0.192,\ -0.123)$. Eftersom intervallet ej täcker över 0 förkastas $H_0 : \beta_1 = 0$ (tiden påverkar ej). Ja, det finns en signifikant förändring av bensenhalt med avseende på tiden.

   (c) Sökt är ett konfidensintervall då $t_0 = 15$.

   Ett 95% konfidensintervall är $(\hat{\beta}_0 + \hat{\beta}_1 \cdot t_0 \pm t_{0.975,12} \sqrt{\hat{\sigma^2}(\frac{1}{11} + \frac{(x_0 - \bar{t})^2}{SS_t}}) =$

   $(3.237 - (-0.1576) \cdot 19 \pm 2.18 \cdot 0.2369 \sqrt{(\frac{1}{14} + \frac{(15-7.5)^2}{227.5}})) = (0.6,\ 1.2)$.

Matematisk statistik
Lunds universitet

Tentamen 2014-03-21 kl 8^{00}–13^{00}
MASB11 — Biostatistisk grundkurs 7.5 hp

Väl motiverade lösningar med svar krävs på varje uppgift, t.ex. ska modeller, approximationer, hypoteser och slutsatser anges och motiveras.

Institutionens papper används både som kladdpapper och som inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas. Skriv fullständigt namn på alla papper.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, Formelsamling "MASB11 Biostatistisk grundkurs" samt miniräknare.

Totalt kan man få 100 poäng. Gränsen för godkänd är 50 poäng och för väl godkänd 75 poäng.

Resultatet läggs senast 2014-03-28 in i ladok.

---

1. (a) Vikten hos fullgångna nyfödda barn anses vara normalfördelad med väntevärde 3.5 kg och standardavvikelse 0.5 kg. Barn med en födelsevikt under 2.5 kg eller över 4.5 kg anses av vissa forskare ha en större risk för plötslig spädbarnsdöd. Hur stor andel av de fullgångna nyfödda barnen ligger i denna riskgrupp? (5p)

   (b) Antalet olyckor per år på en arbetsplats är poissonfördelat med väntevärde tre. Vad är sannolikheten att det förekommer minst sju olyckor under ett år? (5p)

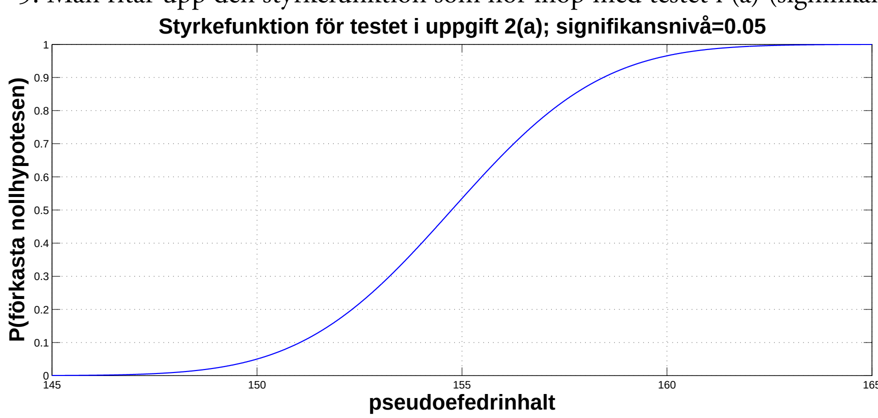
   (c) Fortsättning från 1b. På arbetsplatsen inträffade det ett år sju olyckor. Ställ upp lämpliga hypoteser och testa på signifikansnivå 5 % om detta tyder på att väntevärdet för antalet årliga olyckor är större än 3. (5p)

   (d) I en population har 25 % hund, 20 % katt medan 10 % har både hund och katt. Vi väljer en person slumpmässigt ur populationen, vad är sannolikheten att personen varken har katt eller hund? (5p)

2. Enligt antidopningsbyrån Wada är en koncentration som överstiger 150 mikrogram per milliliter av pseudoefedrin i urinen att betrakta som dopning. På en manlig ishockeyspelare gjordes tre mätningar: 160, 155, 150 (mikrogram per milliliter). Antag att normalfördelning är en rimlig modell.

   (a) Ställ upp lämpliga hypoteser och undersök om data tyder på att ishockeyspelaren är dopad enligt Wadas regler? (12p)

   (b) Man tror sig veta att mätningarna följer en normalfördelning med väntevärde μ och standardavvikelse σ där $\sigma = 5$. Man ritar upp den styrkefunktion som hör ihop med testet i (a) (signifikansnivå 5 %):



   Antag att ishockeyspelaren är dopad så att hans verkliga μ är 155 mikrogram per milliliter. Hur stor är då sannolikheten att han inte åker fast i dopningskontrollen? (8p)

3. Ankyloserande spondylit (AS) leder typiskt till förbeningar i bäckenleder och i ryggens kotpelare. På ett antal gravida kvinnor, med och utan denna diagnos, undersökte man hur många av förlossningarna som slutade i akut kejsarsnitt:

	AS	Ej AS
Kejsarsnitt	36	36
Ej kejsarsnitt	64	164

(a) Verkar det finnas ett samband mellan AS och kejsarsnitt? (10p)

(b) Gör ett konfidensintervall för andelen graviditeter i AS-gruppen som slutar med kejsarsnitt. (7p)

(c) Gör ett konfidensintervall för förväntade antalet graviditeter, som slutar med kejsarsnitt, i en AS-grupp bestående av 200 kvinnor. (3p)

4. I en studie ville man undersöka om en låg dos av aspirin påverkar blodtrycket hos gravida kvinnor som fått högt blodtryck under graviditeten. Ett slumpmässigt urval av 23 kvinnor fick en medicin med aspirin medan 24 andra slumpmässigt utvalda kvinnor fick ett placebo. Efter en tids medicinering mättes blodtrycket hos samtliga. Resultat (mm Hg):

	medelvärde	standardavvikelse	antal
Aspirin	111	7.8	23
Placebo	109	8.2	24

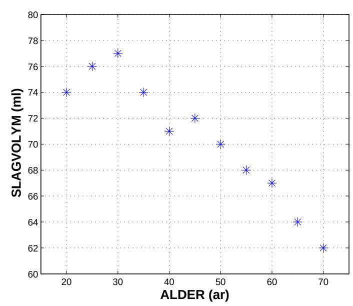
(a) Hur stor är skillnaden i förväntat blodtryck mellan de två grupperna? Gör ett konfidensintervall. Du kan anta att blodtrycket är approximativt normalfördelat i de två grupperna. (10p)

(b) Undersök om det förväntade blodtrycket skiljer sig åt i de två grupperna. (5p)

(c) Kommentera kring studiens försöksupplägg, kan du föreslå ett bättre upplägg när man vill undersöka om aspirin påverkar blodtrycket? (5p)

5. Volymen av blod som strömmar från hjärtat efter en hjärtmuskelsammandragning kallas slagvolymen. Vid en medicinisk undersökning av sambandet mellan slagvolym och ålder fick man följande data:

Ålder:	20	25	30	35	40	45	50	55	60	65	70
Slagvolym:	74	76	77	74	71	72	70	68	67	64	62



Räknehjälp: Om x_i är ålder för person i och y_i slagvolymen för person i gäller att

$$\bar{x} = 45; \quad \bar{y} = 70.4545; \quad s_x^2 = \frac{1}{11-1}\sum(x_i - \bar{x})^2 = 275; \quad s_y^2 = \frac{1}{11-1}\sum(y_i - \bar{y})^2 = 23.2727.$$

Dessutom är

$$SS_x = \sum(x_i - \bar{x})^2 = 2750; \quad SP_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = -755; \quad SS_y = \sum(y_i - \bar{y})^2 = 232.7273.$$

(a) Har åldern en signifikant inverkan på slagvolymen? (6p)

(b) Hur stor är den genomsnittliga skillnaden i slagvolym mellan en 25-åring och en 55-åring? Gör ett lämpligt intervall. (4p)

(c) Vad är den förväntade slagvolymen hos en 50-åring? Gör ett lämpligt intervall. (6p)

(d) För en individ är slagvolymen cirka 75 ml. Gör en uppskattning av individens ålder. (4p)

Lycka till!

Matematisk statistik
Lunds universitet

Kortfattade lösningar till tentamen i
MASB11-Biostatistisk grundkurs,
2014-03-21

1. (a) X = födelsevikt; $X \in N(3.5,\ 0.5^2)$.
$P(X > 4.5) = 1 - P(Z \leq \frac{4.5-3.5}{0.5}) = 1 - P(Z \leq 2) = 1 - 0.9772 = 0.0228$.
$P(X < 2.5) = P(Z \leq \frac{2.5-3.5}{0.5}) = P(Z \leq -2) = 1 - P(Z \leq 2) = 1 - 0.9772 = 0.0228$.
P(barn i riskgruppen)=$2 \cdot 0.0228 = 0.045$.

(b) X = antalet olyckor; $X \in Po(3)$. $P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9665 = 0.0335$.

(c) X = antalet olyckor; $X \in Po(\mu)$. $H_0 : \mu = 3$; $H_1 : \mu > 3$. Testets P-värde (probvärde) är $P = P(X \geq 7) = 0.0335$ (från 1(b)). Eftersom $P < 0.05$ förkastas H_0 på nivå 0.05. Ja, data tyder på att förväntat antal olyckor ökat.

(d) P(hund)=0.25, P(katt)=0.20, P(hund$\cap$katt)=0.10.
P(ej hund $\cap$ ej katt)=1$-$P(katt $\cup$ hund)=1$-$[P(katt)+P(hund)$-$P(hund$\cap$katt)]=
1$-$(0.20+0.25$-$0.10)=1$-$0.35=0.65.

2. (a) X = halten pseudoefedrin; $X \in N(\mu,\ \sigma^2)$.
Intressanta hypoteser är $H_0 : \mu \leq 150$; $H_1 : \mu > 150$.
Från data: $\hat{\mu} = \bar{x} = 155$; $\hat{\sigma} = s = 5$.
Ett ensidigt, undre begränsat, 95% intervall för μ fås genom $I_\mu = (\bar{x} - t_{0.95,2}\frac{s}{\sqrt{3}},\ \infty) = (155-2.92\frac{5}{\sqrt{3}},\ \infty)) = (146.57,\ \infty))$. Eftersom intervallet täcker över 150 kan H_0 ej förkastas på nivå 5 %. Vi har alltså inte påvisat att ishockeyspelaren är dopad.

(b) Styrkefunktionen anger P(H_0 förkastas då μ är den verkliga halten)=P(personen åker fast då μ är den verkliga halten). Genom att avläsa funktionens värde vid en halt på 155 fås P(han åker inte fast då $\mu = 155$)=1-P(han åker fast då $\mu = 155$)$\approx$1-0.53=0.47.

3. (a) Modell: X=blodtryck för en kvinna som får aspirin; $X \in N(\mu_1,\ \sigma_1^2)$; Y=blodtryck för en kvinna som får placebo; $Y \in N(\mu_2,\ \sigma_2^2)$. Eftersom skattningarna av σ_1 och σ_2 är ungefär lika stora antar vi att variansen är den samma i båda grupperna (kan testas) och vi gör en gemansam skattning av variansen:
$\hat{\sigma^2} = \frac{(23-1)\cdot 7.8^2+(24-1)\cdot 8.2^2}{(23-1)+(24-1)} = 64.11$.
Ett 95 % intervall för skillnaden $\mu_1 - \mu_2$ fås genom $I_{\mu_1-\mu_2} = (\bar{x}-\bar{y} \pm t_{0.975,45} \cdot \sqrt{\hat{\sigma}^2(\frac{1}{23} + \frac{1}{24})}) = (111 - 109 \pm 2.02 \cdot \sqrt{64.11(\frac{1}{23} + \frac{1}{24})}) = (-2.72,\ 6.72)$.

(b) Vi vill testa hypotesen $H_0 : \mu_1 - \mu_2 = 0$ mot $H_1 : \mu_1 - \mu_2 \neq 0$.
Eftersom intervallet från (a) täcker över 0 kan H_0 ej förkastas på nivå 0.05. Vi har inte påvisat att det förväntade blodtrycket skiljer sig åt i de två grupperna.

(c) Upplägget är olämpligt eftersom en eventuell skillnad **mellan** grupperna kan vara svår att upptäcka på grund av en stor variation i blodtryck hos kvinnor **inom** grupperna.
Ett bättre försöksupplägg hade varit att låta varje kvinna under en period få placebo och under en annan period få aspirin. Blodtrycket mäts under båda perioderna. På så sätt undviker man

variationer i blodtryck mellan kvinnor och kan mäta det intressanta, nämligen skillnad mellan medlen. Detta försöksupplägg analyseras lämpligen enligt modellen ”stickprov i par”.

4. (a) χ^2-test: Observerade värden är:

	AS	Ej AS	
Kejsarsnitt	36	36	72
Ej kejsarsnitt	64	164	228
	100	200	300

Förväntade värden om H_0:”inget samband” är sann:

	AS	Ej AS
Kejsarsnitt	$\frac{100 \cdot 72}{300} = 24$	$\frac{200 \cdot 72}{300} = 48$
Ej kejsarsnitt	$\frac{100 \cdot 228}{300} = 76$	$\frac{200 \cdot 228}{300} = 152$

$\chi^2 = \frac{(36-24)^2}{24} + \frac{(36-48)^2}{48} + \frac{(64-76)^2}{76} + \frac{(164-152)^2}{152} = 11.48$
Eftersom $\chi^2 = 11.48 > X^2_{0.999,1} = 10.828$ förkastas H_0. Ja, det finns ett signifikant samband.

(b) Modell: X=antal kvinnor i AS-gruppen vars graviditet slutar med kejsarsnitt; $X \in Bin(100, p)$. Sökt är I_p och $\hat{p} = \frac{36}{100}$. Eftersom $n \cdot p \cdot (1-p) \approx 100 \cdot \frac{36}{100} \cdot (1 - \frac{36}{100}) > 10$ är normalapproximation tillåten och ett konfidensintervall för p med den approximativa konfidensgraden 0.95 ges av $I_p = (\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{100}}) = (0.266,\ 0.454)$

(c) Sökt är ett intevall för $200 \cdot p$ vilket blir $(200 \cdot 0.266,\ 200 \cdot 0.454) = (53.18,\ 90.82)$.

5. Modellen är att slagvolymen (y) beskrivs linjärt av åldern (x): $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, där ε_i är oberoende och $N(0,\ \sigma^2)$.

(a) $\hat{\beta}_1 = \frac{SP_{xy}}{SS_x} = \frac{-755}{2750} = -0.2745$. Vi vill göra ett intervall för β_1. Vi behöver först skatta σ: $\hat{\sigma^2} = \frac{1}{11-2}(SS_y - \frac{SP_{xy}^2}{SS_x}) = \frac{25.45}{11-2} = 2.8278$.

Ett 95% intervall för β_1 fås genom $I_{\beta_1} = (\hat{\beta}_1 \pm t_{0.975,9}\frac{\hat{\sigma^2}}{SS_x}) = (-0.3471,\ -0.202)$. Eftersom intervallet ej täcker över 0 förkastas $H_0 : \beta_1 = 0$ (åldern påverkar ej). Ja, åldern har en signifikant påverkan på slagvolymen.

(b) Eftersom β_1 tolkas som hur mycket slagvolymen i genomsnitt ändras då man blir ett år äldre söks här ett intervall för $30\beta_1$. Detta fås genom $I_{30\beta_1} = 30 \cdot I_{\beta_1} =$
$(30 \cdot (-0.3471),\ 30 \cdot (-0.202)) = (-10.413,\ -6.060)$.

(c) Sökt är ett konfidensintervall då $x_0 = 50$. Vi behöver först skatta β_0: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 82.81$.

Ett 95% konfidensintervall är $(\hat{\beta}_0 + \hat{\beta}_1 \cdot x_0 \pm t_{0.975,9}\sqrt{\hat{\sigma^2}(\frac{1}{11} + \frac{(x_0-\bar{x})^2}{SS_x}}) =$
$(82.81 - 0.2745 \cdot 50 \pm 2.26\sqrt{2.8278(\frac{1}{11} + \frac{(50-45)^2}{27.50}})) = (67.87,\ 70.28)$.

(d) Här vet vi att $y = 75$, vilken ålder x motsvarar det? Lös ut x från den skattade regressionslinjen: $x = \frac{y-\hat{\beta}_0}{\hat{\beta}_1} = \frac{75-82.81}{-0.2745} = 28.45$.

Väl motiverade lösningar med svar krävs på varje uppgift, t.ex. ska modeller, approximationer, hypoteser och slutsatser anges och motiveras.

Institutionens papper används både som kladdpapper och som inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas. Skriv fullständigt namn på alla papper.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, Formelsamling ”MASB11 Biostatistisk grundkurs” samt miniräknare.

Totalt kan man få 100 poäng. Gränsen för godkänd är 50 poäng och för väl godkänd 75 poäng.

Resultatet läggs senast 2014-04-11 in i ladok.

---

1. (a) I en högstadieskola är 45% tjejer. Från en enkätundersökning har man informationen att 35% av tjejerna använder cykelhjälm, medan motsvarande siffra för killarna är 30%. Vi väljer ut en elev slumpmässigt, vad är sannolikheten att vi valt en person som inte använder cykelhjälm? (5p)

   (b) Vikten hos en alpin skidåkare med utrustning anses normalfördelad med väntevärde 80 kg och varians 36 kg^2. Skidåkaren Eva åker ensam i kabinen. Vad är sannolikheten att hennes vikt överstiger 90 kg? (5p)

   (c) Vikten hos en alpin skidåkare med utrustning anses normalfördelad med väntevärde 80 kg och varians 36 kg^2. Skidåkarna Eva och Johan åker i kabinen tillsammans. Vad är sannolikheten att deras totala vikt överstiger 190 kg? (5p)

   (d) Dygnsmedeltemperaturen i juli i Luleå varierar år från år enligt en stokastisk variabel X med väntevärde 18 och standardavvikelse 2, där temperaturen är angiven i $^\circ C$. Nu vill man räkna om ovanstående värden till $^\circ F$. Om X är temperaturen i $^\circ C$ så blir $Y = \frac{9 \cdot X}{5} + 32$ motsvarande temperatur i $^\circ F$. Bestäm väntevärde och standardavvikelse för Y. (5p)

2. För att undersöka om en ny typ av foder för boskap ökar tillväxten mäter man vikten innan och efter två månaders utfodring. Man valde att testa fodret på sju slumpmässigt utvalda kalvar i samma ålder. Värdena som erhölls var vikt (kilo) före och efter utfodring:

Kalv:	1	2	3	4	5	6	7
Före:	300	270	340	328	296	335	290
Efter:	335	314	372	365	338	372	318

   Normal viktökning för en kalv i den givna åldern är 30 kilo över två månader. Mätvärdena för vikten kan antas vara normalfördelade. Undersök om man kan påvisa någon effekt i form av tillväxtökning för denna nya typ av foder med hjälp av ett lämpligt konfidensintervall eller test. (20p)

3. I en undersökning av metaller i biota mätte man ett år halten Cd (mg/kg) i lever och njure på 12 älgar i Kronobergs län. Samtidigt bedömdes älgarnas ålder. Resultat för Cd-halt i lever:

Älg nr	1	2	3	4	5	6	7	8	9	10	11	12
Ålder (år)	0.5	1.5	0.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	0.5	1.5
Cd-halt (mg/kg)	0.15	0.69	0.37	0.51	0.30	0.22	0.56	0.96	1.10	0.25	0.19	0.33

   Hur stor är skillnaden i förväntad Cd-halt i lever mellan älgar av ålder 0.5 år och 1.5 år? Svara i form av ett lämpligt konfidensintervall. Du får anta lämplig(a) normalfördelningar. (20p)

4. I England har under en längre period födelsetalen bland gravida kvinnor varit att det föds 106 pojkar på 100 flickor. I en större undersökning specialstuderade man 386 gravida kvinnor som var vegetarianer och fann att bland dessa 386 kvinnor föddes det 180 pojkar. Man vill undersöka om data från den gjorda undersökningen tyder på att vegetarianer föder färre pojkar.

(a) Ange en modell för data och ställ upp lämpliga hypoteser. Bortse från tvilling- och trillingfödslar. (7p)

(b) Undersök om data från den gjorda undersökningen tyder på att vegetarianer föder färre pojkar genom att beräkna ett lämpligt intervall eller utföra ett lämpligt test. (8p)

(c) Antag att vi väljer slumpmässigt ut två gravida vegetarianer i England. Gör en skattning av sannolikheten att precis en av dem får en pojke. (5p)

5. Bensen sprids i luften från bland annat bensindrivna fordon. Efter att ha infört katalytisk avgasrening och striktare regler för hur mycket bensen som får användas i bensinen har bensenhalten i luft, μg/m^3, i tätortsmiljö sjunkit sedan början av 1990-talet. I nedanstående tabell från Naturvårdsverket finns uppmätt bensenhalt i en svensk medeltätort för perioden 1994–2007, ($t_1 = 1$ motsvarar år 1994).

t_i	1	2	3	4	5	6	7
Bensen y_i (μg/m^3)	3.47	2.79	2.93	2.28	2.08	2.44	1.92

t_i	8	9	10	11	12	13	14
Bensen y_i (μg/m^3)	1.95	2.02	1.94	1.41	1.22	1.33	0.99

Räknehjälp:

$\sum t_i = 105$; $\sum y_i = 28.77$;
$SS_t = \sum(t_i - \bar{t})^2 = 227.5$;
$SP_{ty} = \sum(t_i - \bar{t})(y_i - \bar{y}) = -35.855$;
$SS_y = \sum(y_i - \bar{y})^2 = 6.3244$.

(a) Ansätt en linjär regressionsmodell $y_i = \beta_0 + \beta_1 \cdot t_i + \varepsilon_i$, $\varepsilon_i \in N\left(0,\ \sigma^2\right)$, för hur bensenhalten varierar med tiden och skatta alla tre parametrarna i modellen. (7p)

(b) Avgör på lämpligt sätt om förändringen av bensenhalten med avseende på tiden är signifikant. (6p)

(c) Utifrån dessa data, vad är den förväntade bensenhalten för år 2008? Svara med ett lämpligt intervall. (7p)

Lycka till!

Matematisk statistik
Lunds universitet

Kortfattade lösningar till tentamen i
MASB11-Biostatistisk grundkurs,
2014-04-04

1. (a) P(cykelhjälm)=P(tjej)·P(cykelhjälm|tjej)+P(kille)·P(cykelhjälm|kille)=$0.45\cdot 0.35+0.55\cdot 0.3 = 0.3225$. P(ej cykelhjälm)$= 1-0.3225 = 0.6775$.
   (b) X=Evas vikt; $X \sim N(80,\ 36)$. $P(X > 90) = 1 - P(Z \leq \frac{90-80}{\sqrt{36}}) = 1 - 0.952 = 0.048$.
   (c) $X_1 + X_2$=total vikt; $X_1 + X_2 \sim N(80+80,\ 36+36)$.
   $P(X_1 + X_2 > 190) = 1 - P(Z \leq \frac{190-160}{\sqrt{2\cdot 36}}) = 2.03 \cdot 10^{-4}$.
   (d) $E(Y) = E(\frac{9\cdot X}{5} + 32) = \frac{9\cdot E(X)}{5} + 32 = \frac{9\cdot 18}{5} + 32 = 64.4$.
   $V(Y) = V(\frac{9\cdot X}{5} + 32) = \frac{9^2\cdot V(X)}{5^2} = \frac{9^2\cdot 2^2}{5^2}$.
   $\sqrt{V(Y)} = \frac{9\cdot 2}{5} = \frac{18}{5} = 3.6$.

2. Modell (stickprov i par): Z=viktökning; $Z \sim N(\mu_d,\ \sigma^2)$. Vi har 7 observationer $z_1, \ldots, z_7$ från Z där z_i är "vikt efter" - "vikt före" för kalv i. De 8 z-värdena är $35, 44, 32, 37, 42, 37, 28$. Medelvärdet för differenserna är $\bar{z} = 36.4286$ och skattad standardavvikelse för differenserna är $s_z = 5.5032$. Intressanta hypoteser är:
   $H_0 : \mu_d = 30$, (fodret ger normal viktökning)
   $H_1 : \mu_d > 30$ (fodret ger tillväxtökning).
   Gör ett ensidigt undre begränsat 95% konfidensintervall för μ_d: $I_{\mu_d} = (\bar{z} - t_{0.95,6}\frac{s_z}{\sqrt{7}}, \infty) = (32.4,\ \infty)$. Eftersom intervallet ej täcker över 30 drar vi slutsatsen att H_0 förkastas på nivå 5%. Ja, det finns en statistisk påvisbar tillväxtökning med detta foder.

3. Modell:
   Cd-halt från älgar med ålder 0.5 år, $x_1, \ldots, x_7$ observationer av $X \sim N(\mu_1,\ \sigma^2)$
   Cd-halt från älgar med ålder 1.5 år, $y_1, \ldots, y_5$ observationer av $Y \sim N(\mu_2,\ \sigma^2)$
   Med hjälp av räknare fås
   $\bar{x} = 0.2843,\quad s_x = 0.123$
   $\bar{y} = 0.7280,\quad s_y = 0.3083$
   Eftersom de två σ-skattningarna är någorlunda lika anser vi att antagandet om lika σ^2 i modellen är OK och vi gör en gemensam skattning av σ^2:
   $\hat{\sigma} = s_p = \sqrt{\frac{(7-1)s_x^2+(5-1)s_y^2}{(7-1)+(5-1)}} = 0.2170$
   Lämpliga hypoteser:
   $H_0 : \mu_1 = \mu_2$, dvs ingen skillnad i genomsnittlig Cd-halt
   $H_0 : \mu_1 \neq \mu_2$, dvs skillnad i genomsnittlig Cd-halt
   Gör ett konfidensintervall för $\mu_1 - \mu_2$:
   $I_{\mu_1-\mu_2} = (\bar{x} - \bar{y} \pm t_{0.975,10} s_p\sqrt{\frac{1}{7} + \frac{1}{5}}) = (0.2843 - 0.7280 \pm 2.23 \cdot 0.2170\sqrt{\frac{1}{7} + \frac{1}{5}}) = (-0.7269,\ -0.1606)$.

Eftersom intervallet ej täcker över 0, kan H_0 förkastas på nivå 0.05. Det finns en skillnad mellan förväntad Cd-halt mellan åldersgrupperna. Älgar med ålder 1.5 år har i genomsnitt mellan $(0.16, 0.73)$ enheter högre Cd-halt i lever än älgar av ålder 0.5 år.

4. (a) Låt X=antalet pojkar som de 386 kvinnliga vegetarianerna föder. Då är $X \sim Bin(386,\ p)$, där $p =$P(vegetarian föder pojke). För befolkningen i gemen gäller att motsvarande sannolikhet är $p_0 = \frac{106}{106+100} = 0.5146$. De intressanta hypoteserna är
   $H_0 : p = 0.5146$,
   $H_1 : p < 0.5146$.

   (b) Enklast är att använda direktmetoden och beräkna prob-värdet. Vi observerade att i undersökningen föddes 180 pojkar i vegetariangruppen.

   prob-värde=P($X \leq 180$ om $X \sim Bin(386,\ 0.5146)) = 0.03$, där sannolikheten är beräknad m.h.a. normalapproximation:

   Eftersom prob-värdet understiger "standardgränsen" 0.05 kan H_0 förkastas på nivå 0.05. I själva verket kan den förkastas på nivå 0.03. Ja, det tycks alltså att kvinnliga vegetarianer föder färre pojkar.

   (c) $\hat{p} = \frac{180}{386}$. P(precis en av dem får en pojke)=2$\cdot$ P(den ena får en pojke och den andra får en flicka)$\approx 2 \cdot \frac{180}{386} \cdot (1 - \frac{180}{386}) = 0.4977$.

5. (a) $\hat{\beta}_1 = \frac{SP_{ty}}{SS_t} = \frac{-35.855}{227.5} = -0.1576$.
   $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{t} = 3.237$
   $\hat{\sigma} = \sqrt{\frac{1}{14-2}(SS_y - \frac{SP_{ty}^2}{SS_t})} = \sqrt{\frac{0.6734}{14-2}} = 0.2369$.

   (b) Ett 95% intervall för β_1 fås genom $I_{\beta_1} = (\hat{\beta}_1 \pm t_{0.975,12}\frac{\hat{\sigma}}{\sqrt{SS_t}}) = (-0.192,\ -0.123)$. Eftersom intervallet ej täcker över 0 förkastas $H_0 : \beta_1 = 0$ (tiden påverkar ej). Ja, det finns en signifikant förändring av bensenhalt med avseende på tiden.

   (c) Sökt är ett konfidensintervall då $t_0 = 15$.

   Ett 95% konfidensintervall är $(\hat{\beta}_0 + \hat{\beta}_1 \cdot t_0 \pm t_{0.975,12}\sqrt{\hat{\sigma}^2(\frac{1}{14} + \frac{(x_0-\bar{t})^2}{SS_t})} =$

   $(3.237 + (-0.1576) \cdot 15 \pm 2.18 \cdot 0.2369\sqrt{(\frac{1}{14} + \frac{(15-7.5)^2}{227.5}})) = (0.6,\ 1.2)$.

Väl motiverade lösningar med svar krävs på varje uppgift, t.ex. ska modeller, approximationer, hypoteser och slutsatser anges och motiveras.

Institutionens papper används både som kladdpapper och som inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas. Skriv fullständigt namn på alla papper.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, Formelsamling "MASB11 Biostatistisk grundkurs" samt miniräknare.

Totalt kan man få 100 poäng. Gränsen för godkänd är 50 poäng och för väl godkänd 75 poäng.

Resultatet läggs senast 2015-06-12 in i ladok.

---

1. (a) Tiden det tar för Olga att cykla mellan hem och jobb är normalfördelad med väntevärde 15 (min) och standardavvikelse 2 (min). Olga börjar jobbet 8.30, vad är sannolikheten att hon hinner i tid om hon startar cykla 8.13? (5p)

   (b) I en population anger 40 % att de motionerar regelbundet och 60 % att de har ett stillasittande jobb. Det är 20 % som både har ett stillasittande jobb och motionerar regelbundet. Beräkna sannolikheten att, en person som har ett stillasittande jobb, inte motionerar regelbundet. Ledning: Att rita ett Venndiagram kan vara till hjälp. (5p)

   (c) Tillfällen då ett sjukhus drabbas av akut blodbrist kan modelleras enligt en poissonprocess med intensitet $\lambda = 0.5$ (per år). Det innebär att antal tillfällen med blodbrist under ett år är poissonfördelat med väntevärde 0.5. Beräkna sannolikheten att man under en fyraårsperiod får högst ett tillfälle med blodbrist. (5p)

   (d) Vikten hos en 9-åring anses vara normalfördelad med väntevärde 30 kg och standardavvikelse 3 kg. Två slumpmässigt utvalda 9-åringar ska tillsammans bära en låda som väger 10 kg över en hängbro. Beräkna väntevärde och standardavvikelse för den totala vikten på bron när barnen går över. (5p)

2. Ett antal personer ingick i en studie där man ville undersöka om en mycket uppmärksammad diet sänker det systoliska blodtrycket. Försökspersoner med ett "måttligt" förhöjt blodtryck undersöktes vid starten av studien och sedan efter 6 månaders diet. Person nr 4 fick på grund av sjukdom avbryta sitt deltagande i studien.

Person nr	1	2	3	4	5	6	7
Blodtryck före (mmHg)	150	160	145	165	165	155	150
Blodtryck efter (mmHg)	145	155	150	-	160	145	140

   (a) Tyder dessa data på att dieten sänker blodtrycket? Sätt upp lämpliga hypoteser och utför ett test. Lämplig(a) normalfördelningar får antas. (15p)

   (b) "Dieten har störst effekt på de som är sjukast". För att undersöka detta påstående görs motsvarande studie på sex personer med allvarlig blodtrycksförhöjning. Lämplig(a) normalfördelningar får antas.

Person	8	9	10	11	12	13
Före	170	185	175	190	175	195
Efter	155	160	165	160	150	170

   Beskriv (ange t.ex. modell och hypoteser) hur du skulle gå till väga för att pröva om påståendet verkar stämma. Du behöver inte göra en fullständig beräkning i denna deluppgift. (5p)

3. Hyperhidros innebär att en person har onormalt mycket svettningar. På 20 personer som samtliga medicinerade mot åkomman mätte man svettmängd (mg/min).

Person nr	1	2	3	4	5	6	7	8	9	10
Svettmängd (mg/min)	29	91	4	20	9	23	63	76	12	21

Person nr	11	12	13	14	15	16	17	18	19	20
Svettmängd (mg/min)	38	25	45	50	13	28	4	104	18	10

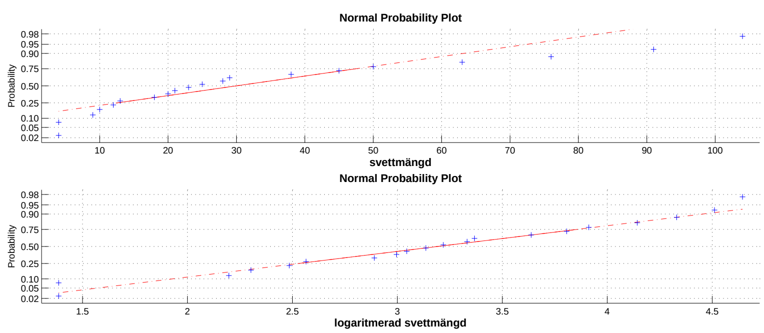
Räknehjälp: Om x_i är svettmängd hos person i gäller $\bar{x} = 34.15; s_x = \sqrt{\frac{1}{20-1}\sum_{i=1}^{20}(x_i - \bar{x})^2} = 28.9796;$. Om $y_i = \ln(x_i)$ är logaritmerad svettmängd hos person i gäller $\bar{y} = 3.1646; s_y = \sqrt{\frac{1}{20-1}\sum_{i=1}^{20}(y_i - \bar{y})^2} = \sqrt{\frac{1}{20-1}\sum_{i=1}^{20}(\ln(x_i) - \overline{\ln(x)})^2} = 0.9328.$

För att undersöka om data kunde modelleras med en normalfördelning ritades de ut i ett normalfördelningspapper (övre delen av figuren nedan) som fungerar i princip som en QQ-plot. För att undersöka om en lognormalfördelning var en bättre modell ritade man också ut logaritmerad svetthalt (undre del av figuren) i samma typ av papper.

(a) Vilken av de två modellerna passar bäst till data? Motivera ditt svar. (3p)

(b) Utnyttja den fördelning du valt i (a) och beräkna sannolikheten att svettmängden överstiger 100 mg/min. Använd gärna räknehjälpen under tabellen. (10p)

(c) Antag att vi har 10 patienter som använder den aktuella medicinen. Beräkna sannolikheten att minst en av dem kommer ha svettningar som överstiger 100 mg/min. (7p)



4. Man ville undersöka om promillehalten i blodet påverkas av vikten. Nio kvinnor drack en starköl under en timme varefter deras promillehalt i blodet mättes:

Vikt (kg)	63	86	71	59	79	65	54	62	58
Halt (promille)	0.15	0.07	0.13	0.18	0.10	0.16	0.17	0.16	0.17

Modell: Man antar att halten beror linjärt på vikten bortsett från en slumpmässig avvikelse som är normalfördelad med varians σ^2.

Räknehjälp: $\bar{x} = 66.3333$; $\bar{y} = 0.1433$; $SS_x = 876$; $SS_y = 0.0108$; $SP_{xy} = -2.99$

(a) Undersök om vikten påverkar promillehalten. (8p)

(b) Pia väger 55 kg medan Petra väger 68 kg. Gör ett 95% konfidensintervall för den förväntade skillnaden i deras promillehalt då de båda druckit en öl under en timme. (4p)

(c) Gör ett 95% konfidensintervall för den förväntade promillehalten hos en kvinna som väger 55 kg och som druckit en öl under en timme. (8p)

5. Av de 675 kvinnliga patienter som registrerats för akut hjärtinfarkt vid ett större sjukhus dog 49 av sin infarkt. Under samma period registrerades 1056 manliga patienter för samma åkomma varav 55 dog.

(a) Gör ett approximativt 95 % intervall för andelen kvinnor som dör av sin akuta infarkt. (10p)

(b) Tyder dessa data på att det finns någon skillnad mellan könen beträffande överlevnad i akut infarkt? (10p)

Lycka till!

Matematisk statistik
Lunds universitet

Kortfattade lösningar till tentamen i
MASB11-Biostatistisk grundkurs,
2015-06-01

1. (a) X=cykeltid; $X \sim N(15,\ 2^2)$. P(Olga i tid)=P($X \leq 17$)=$\Phi(\frac{17-15}{2}) = \Phi(1) = 0.841$.
   (b) P(ej motion| stilla jobb)=P(ej motion $\cap$ stilla jobb) / P(stilla jobb)= $\frac{0.6-0.2}{0.6} = \frac{2}{3}$. Rita gärna ett Venn-diagram.
   (c) X antal perioder med blodbrist under en fyraårsperiod; $X \sim Po(4 \cdot 0.5)$; $P(X \leq 1) = 0.406$.
   (d) X_i=vikt hos barn nr i, $i = 1, 2$. Total vikt på bron är $X_1 + X_2 + 10$. $E(X_1 + X_2 + 10) = 30+30+10 = 70$ kg; $V(X_1+X_2+10) = V(X_1)+V(X_2) = 9+9 = 18$. $D(X_1+X_2+10) = \sqrt{V(X_1 + X_2 + 10)} = \sqrt{18}$ kg.

2. (a) Modell (stickprov i par): Z=blodtrycksförändring (måttlig förhöjning); $Z \sim N(\mu_d,\ \sigma^2)$. Vi har 6 observationer (person nr 4 stryks) $z_1, \ldots, z_6$ från Z där z_i är "blodtryck före" - "blodtryck efter" för person i. De 6 z-värdena är $5, 5, -5, 5, 10, 10$. Medelvärdet för differenserna är $\bar{z} = 5$ och skattad standardavvikelse för differenserna är $s_z = 5.4772$. Intressanta hypoteser:
   $H_0 : \mu_d = 0$, (ingen blodtryckdförändring)
   $H_1 : \mu_d > 0$ (sänkning av blodtryck).
   Gör ett ensidigt undre begränsat 95% konfidensintervall för μ_d: $I_{\mu_d} = (\bar{z} - t_{0.95,5}\frac{s_z}{\sqrt{6}},\ \infty) = (0.494,\ \infty)$. Eftersom intervallet ej täcker över 0 drar vi slutsatsen att H_0 förkastas på nivå 5%. Ja, det finns en statistisk påvisbar blodtrycksminskning med denna diet.
   (b) Modell (två oberoende stickprov på förändring i blodtryck): V=blodtrycksförändring (allvarlig förhöjning); $V \sim N(\mu_{d2},\ \sigma^2)$. Vi har 6 observationer $v_1, \ldots, v_6$ från V där v_i är "blodtryck före" - "blodtryck efter" för person i. De 6 v-värdena är $15, 25, 10, 30, 25, 25$. Medelvärdet för differenserna är $\bar{v} = 21.6667$ och skattad standardavvikelse för differenserna är $s_v = 7.5277$.
   Intressanta hypoteser är:
   $H_0 : \mu_d = \mu_{d2}$, (samma blodtrycksförändring)
   $H_1 : \mu_d < \mu_{d2}$ (större blodtrycksförändring för de med högre blodtryck).
   En gemensam skattning av σ fås genom $\hat{\sigma} = s_p = \sqrt{\frac{(6-1)s_z^2+(6-1)s_v^2}{(6-1)+(6-1)}} = 6.5828$
   Gör ett uppåt begränsat konfidensintervall för $\mu_d - \mu_{d2}$:
   $I_{\mu_d-\mu_{d2}} = (-\infty,\ \bar{z}-\bar{v}+t_{0.95,10}s_p\sqrt{\frac{1}{6}+\frac{1}{6}}) = (-\infty,\ 5-21.6667\pm 2.23\cdot 6.5828\sqrt{\frac{1}{6}+\frac{1}{6}}) = (-\infty,\ -23.55)$.
   Intervallet täcker ej 0, H_0 förkastas på nivå 0.05, ja dieten verkar ha större effekt (sänker blodtrycket mer) för de med högre systoliskt blodtryck i utgångsläget.

3. (a) Lognormalfördelning är den lämpligaste av de två modellerna eftersom logaritmerade data anpassar sig bättre till en rät linje i ett normalfördelningsdiagram.
   (b) Modell: X=svettmängd; $\ln(X) \sim N(\mu,\ \sigma^2)$, där $\hat{\mu} = \bar{y} = 3.1646$ och $\hat{\sigma} = s_y = 0.9328$.
   $P(X > 100) = P(\ln(X) > \ln(100)) \approx 1 - \Phi(\frac{\ln(100)-3.1646}{0.9328}) = 0.061$.
   (c) Modell: Z=antal patienter av de 10 som har svettmängd över 100 mg/min; $Z \sim Bin(10,\ 0.061)$.
   $P(Z \geq 1) = 1 - P(Z = 0) = 1 - (1 - 0.061)^{10} = 1 - 0.533 = 0.467$.

4. Modell: Om x_i och y_i är vikt respektive promillehalt hos kvinna i gäller $y_i = b_0 + b_1 x_i + \varepsilon_i$, där $\varepsilon_1, \ldots, \varepsilon_9$ är oberoende och $N(0,\ \sigma^2)$.

   (a) $\hat{\beta}_1 = \frac{SP_{xy}}{SS_x} = \frac{-2.99}{876} = -0.003413.$
   $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 0.3697$
   $\hat{\sigma} = \sqrt{\frac{1}{9-2}(SS_y - \frac{SP_{xy}^2}{SS_x})} = \sqrt{\frac{0.0005944}{9-2}} = 0.009215.$
   Ett 95% intervall för β_1 fås genom $I_{\beta_1} = (\hat{\beta}_1 \pm t_{0.975,7} \frac{\hat{\sigma}}{\sqrt{SS_x}}) = (-0.00415,\ -0.00268)$. Eftersom intervallet ej täcker över 0 förkastas $H_0 : b_1 = 0$ (vikt påverkar ej). Ja, vikten påverkar promillehalten.

   (b) Eftersom b_1 tolkas som hur mycket halten i genomsnitt ändras då man blir ett kilo tyngre söks här ett intervall för $(68-55)b_1$. Detta fås genom $I_{13b_1} = 13 \cdot I_{b_1} =$
   $(13 \cdot (-0.00415),\ 13 \cdot (-0.00268)) = (-0.0539,\ -0.0348).$

   (c) Sökt är ett konfidensintervall för linjens läge då $x_0 = 55$.
   Ett 95% konfidensintervall är $(\hat{b}_0 + \hat{b}_1 \cdot x_0 \pm t_{0.975,7}\sqrt{\hat{\sigma}^2(\frac{1}{9} + \frac{(x_0-\bar{x})^2}{SS_x})} =$
   $(0.3697 + (-0.003413) \cdot 55 \pm 2.365 \cdot 0.009215\sqrt{(\frac{1}{9} + \frac{(55-66.3333)^2}{876}}) = (0.17,\ 0.19).$

5. (a) Modell: X=kvinnor som dör av sin akuta infarkt; $X \in Bin(675,\ p_k)$. Sökt är I_{p_k} och $\hat{p}_k = \frac{49}{675}$. Eftersom $n \cdot p_k \cdot (1 - p_k) \approx 100 \cdot \frac{49}{675} \cdot (1 - \frac{49}{675}) > 10$ är normalapproximation tillåten och ett konfidensintervall för p_k med den approximativa konfidensgraden 0.95 ges av $I_{p_k} =$
   $(\hat{p}_k \pm 1.96\sqrt{\frac{\hat{p}_k(1-\hat{p}_k)}{675}}) = (0.053,\ 0.0922).$

   (b) Lösningsalternativ I: Y=män som dör av sin akuta infarkt; $X \in Bin(1056,\ p_m)$, där $\hat{p_m} = \frac{55}{1056}$. Intressanta hypoteser: $H_0 : p_m = p_k$; $H_1 : p_m \neq p_k$. Ett konfidensintervall för $p_k - p_m$ med den approximativa konfidensgraden 0.95 ges av $I_{p_k - p_m} = (\hat{p}_k - \hat{p_m} \pm 1.96\sqrt{\frac{\hat{p}_k(1-\hat{p}_k)}{675} + \frac{\hat{p_m}(1-\hat{p_m})}{1056}}) =$ $(-0.003,\ 0.044)$. Eftersom intervallet täcker över 0, kan H_o inte förkastas på nivå 0.05 och vi har, med dessa data, inte visat en skillnad mellan könen i överlevnad.
   Lösningsalternativ II: X^2-test: Observerade värden är:

	Dog	Överlevde	
Kvinna	49	626	675
Man	55	1001	1056
	104	1627	1731

   Förväntade värden om H_0:"inget samband kön och överlevnadsbenägenhet" är sann:

	Dog	Överlevde
Kvinna	$\frac{675 \cdot 104}{1731} = 40.5546$	$\frac{675 \cdot 1627}{1731} = 634.4454$
Man	$\frac{1056 \cdot 104}{1731} = 63.4454$	$\frac{1056 \cdot 1627}{1731} = 992.5546$

   $\chi^2 = \frac{(49-40.5546)^2}{40.5546} + \frac{(626-634.4454)^2}{634.4454} + \frac{(55-63.4454)^2}{63.4454} + \frac{(1056-992.5546)^2}{992.5546} = 3.067$
   Eftersom $\chi^2 = 3.067 < X^2_{0.95,1} = 3.84$ kan H_0 ej förkastas. Vi har inte, med dessa data, visat att det finns ett signifikant samband.