

Lösningsförslag

UPPGIFT 1

	Kvinna	Man	
Högskoleutbildad	$0,90 \cdot 0,70 = 0,63$	$0,80 \cdot 0,30 = 0,24$	0,87
Ej högskoleutbildad	0,07	0,06	0,13
	0,70	0,30	1,00

- a) $\Pr(\text{ej högskoleutbildad kvinna}) = 0,07 = 7\%$
- b) $\Pr(\text{högskoleutbildad}) = 0,87$
- c) $\Pr(\text{Kvinna}) \cdot \Pr(\text{Högskoleutbildad}) = 0,70 \cdot 0,87 = 0,609$ vilket inte är detsamma som $\Pr(\text{Kvinna och Högskoleutbildad}) = 0,63$.

Alternativt kan man konstatera att sannolikheten för att en trainee är högskoleutbildad ej är densamma som den betingade sannolikheten att en kvinnlig trainee är högskoleutbildad t ex.

UPPGIFT 2

- a) X är Binomialfördelad med $n=6$ och $p=0,5$
- b) $\Pr(X=2) = 0,234 = 23,4\%$
(Den kumulerade sannolikheten för 2 är 34,4%)
- c) X är approximativt NF med $\mu = np = 300 \cdot 0,5 = 150$ och

$$\sigma = \sqrt{np(1-p)} = \sqrt{300 \cdot 0,5 \cdot 0,5} = \sqrt{75} = 8,66$$

$$\text{Sökt: } \Pr(X < 100) = \Pr\left(z < \frac{100 - 150}{8,66}\right) = \Pr(z < -5,77) \approx 0\%$$

UPPGIFT 3

Homogenitetstest!

H_0 : Lamporna är likvärdiga

H_1 : Det finns skillnader mellan de olika märkena

Obs frekv:	A	B	C	D	
Acceptabla	88	92	95	89	364
Oacceptabla	12	8	5	11	36
	100	100	100	100	400

0,91

0,09

Förv frekv:	A	B	C	D	
Acceptabla	91	91	91	91	364
Oacceptabla	9	9	9	9	36
	100	100	100	100	400

		O	E	O-E	$(O-E)^2/E$
Acceptabla	A	88	91	-3	0,098901
	B	92	91	1	0,010989
	C	95	91	4	0,175824
	D	89	91	-2	0,043956
Oacceptabla	A	12	9	3	1
	B	8	9	-1	0,111111
	C	5	9	-4	1,777778
	D	11	9	2	0,444444
SUMMA				0	3,663004

$$df = (c-1)(r-1) = (4-1)(2-1) = 3$$

Observerat Chi-2-värde = 3,66 < 4,642 som är gränsen för 20% signifikans.

Då $P > 20\%$ behålls H_0 . Testresultatet visar inte på någon egentlig skillnad mellan de olika märkena.

Uppgift 4

a)

Slumpvariabeln X får beteckna mängden dricka i ett glas (cl).

Förutsättningar:

X är Normalfördelad med väntevärde = 20cl och standardavvikelsen = 1cl

Slumpmässiga urval där varje stickprov utgörs av 25 observationer

Oberoende mätningar

Fråga: Mellan vilka gränsvärden ligger 90% av stickprovsmedelvärdena?

OBS! Vi studerar stickprovsmedelvärden vilket gör att vi måste känna till vilken sannolikhetsfördelning som dessa följer.

X är normalfördelad med $(\mu_X = 20; \sigma_X = 1)$

vilket innebär att

$\bar{X}$ är normalfördelad med $(\mu_{\bar{X}} = 20; \sigma_{\bar{X}} = \frac{1}{\sqrt{25}})$

Gränsvärdena kommer att hamna:

Enligt z-tabellen ligger 90% av alla värden inom 1,645 standardavvikelser från medelvärdet

Övre gränsvärde: $20 + 1,645 * 0,2 = 20,329$

Undre gränsvärde: $20 - 1,645 * 0,2 = 19,671$

SVAR: Acceptabla medelvärden ligger mellan 19,67 och 20,33 cl

b)

Om medelvärdet ökar till 21 cl så är $\bar{X}$ är normalfördelad med $(\mu_{\bar{X}} = 21; \sigma_{\bar{X}} = \frac{1}{\sqrt{25}})$

Vi kommer att upptäcka denna ökning då medelvärdet överstiger 20,33 (se uppgift a)

Sannolikheten att upptäcka denna ökning är då:

$$P(\bar{X} > 20,329) = 1 - P(\bar{X} < 20,329) = 1 - P\left(Z < \frac{20,329 - 21}{1/\sqrt{25}}\right) = 1 - P(Z < -3,355) = 1 - 0,0004 = 99,96\%$$

SVAR: Sannolikheten att detta upptäcks är nästan 100%

Uppgift 5

a)

SBP = kvantitativ, kontinuerlig, kvotskalnivå

Ålder = kvantitativ, kontinuerlig, kvotskalnivå

Kön = kvalitativ, nominalskalnivå

b)

Lämpliga deskriptiva mått är:

SBP = medelvärde och standardavvikelse, median om materialet är snedfördelat

Ålder = medelvärde och standardavvikelse, median om materialet är snedfördelat

Kön = typvärde

c)

Lämpliga diagram är:

SBP = histogram, dotplot, boxplot

Ålder = histogram, dotplot, boxplot

Kön = stapeldiagram eller cirkeldiagram

d)

Stickprovsmedelvärde $\bar{x} = \frac{\sum x}{n}$

$$\bar{x}_M = \frac{\sum x}{n} = \frac{144 + 220 + \dots + 175}{13} = \frac{2038}{13} = 156,76923$$

$$\bar{x}_K = \frac{\sum x}{n} = \frac{142 + 170 + \dots + 125}{17} = \frac{2258}{17} = 132,82353$$

Uppgift 6

a)

Ja, det ser ut att finnas ett linjärt samband mellan ålder och SBP. Ju äldre man är desto högre tycks värdet på SBP vara. Det finns ett extremvärde som inte passar in (en man med väldigt högt blodtryck), men de övriga observationerna ser ut att passa bra till en tänkt regressionslinje.

b)

Korrelationskoefficienten mellan SBP och ålder är 0,679 vilket är ett tecken på ett relativt starkt linjärt samband. Sambandet är positivt och korrelationskoefficienten är signifikant skild från noll ($p=0,000$) vilket innebär att det finns ett linjärt samband mellan variablerna. Korrelationskoefficienterna som beräknats mellan kön och de andra variablerna är inte relevant att studera eftersom variabeln kön inte är kvantitativ.

c)

Den skattade modellen är signifikant och har en förklaringsgrad på 46%. Dvs 46% av variationen i SBP förklaras av variansen för ålder.

De båda koefficienterna i regressionsmodellen är signifikant skilda från noll.

I genomsnitt så ökar värdet på SBP med ca 1mmHG årligen (1,101).

Det är inte aktuellt att tolka interceptet, skärningen med y-axeln, eftersom en regressionsmodell bara är giltig inom intervallet för observerade värden. Det finns inga observationer med väldigt låga värden på ålder. Den yngste personen var 17 år gammal och de observerade värdena ligger mellan 17-69 år.

d)

Modellen tycks bli bättre eftersom vi får en mindre residualspridning och en högre förklaringsgrad. De båda oberoende variablerna är signifikant skilda från noll och kön tycks alltså påverka sambandet. Vi kan se att kön har ett samband med SBP och att detta är signifikant ($p=0,034$). Om man tar hänsyn till variabeln kön så ökar värdet på SBP i genomsnitt med 0,8mmHg för varje år. När vi tar hänsyn till kön så minskar ålderns inverkan på SBP från 1,101 till 0,831.

Vi kan se att männen tycks ha högre blodtryck än kvinnorna. Männen ligger i genomsnitt 14 mmHg över kvinnorna.