Kapitel 10

10.2

Jämför två guld gruvor genom att ta ett antal sampels från varje och mäter ounce guld/ton malm.

I den första fås $\bar{X}_1 = 0.233$

I den andra fås $\bar{X}_2 = 0.127$

skatta skillnaden $\mu_1 - \mu_2$.

$$\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2 = 0.106$$

10.4

Låt $\bar{X}_1$ vara medel av 25 stickprov
från en $N(8,16)$-fördelning och
$\bar{X}_2$ samma från en $N(5,9)$ med 36 stickprov.

a) Vad är fördelningen för X1bar?

$\bar{X}_1$

Sats 7.3.4: $X_1, \dots, X_n \sim N(\mu, \sigma^2) \Rightarrow N(\mu, \sigma^2/n)$

I vårt fall $X_1, \dots, X_{25} \sim N(8,16) \Rightarrow \bar{X}_1 \sim N(8, 16/25)$

b) Vad är fördelningen för X2bar?

$\bar{X}_2 \sim N(5, 9/36) = N(5, 1/4)$

c)

$(\bar{X}_1 - 8)/(4/5)$

$E(\bar{X}_1) = E(X_1) = 8$

$var(\bar{X}_1) = \frac{16}{25} = \left(\frac{4}{5}\right)^2$

$\Rightarrow \frac{\bar{X}_1 - 8}{4/5} \sim N(0,1)$

d)

$$\frac{\bar{X}_2 - 5}{3/6} ?$$

$$E(\bar{X}_2) = E(X_2) = 5$$

$$var(\bar{X}_2) = \frac{var(X_2)}{n} = \frac{9}{36} = \left(\frac{3}{6}\right)^2$$

$$\Rightarrow \frac{\bar{X}_2 - 5}{3/6} \sim N(0,1)$$

e)

$$\bar{X}_1 - \bar{X}_2 ?$$

Sats: 10.1.1, $\bar{X}_1, \bar{X}_2$ sampels av storlek n_1, n_2 vänte värden μ_1, μ_2 varianser σ_1^2, σ_2^2

$$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(3, \frac{16}{25} + \frac{1}{4}\right)$$

f)

$$\frac{\bar{X}_1 - \bar{X}_2 - (8-5)}{\sqrt{16/25 + 1/4}}$$

$$E(\bar{X}_1 - \bar{X}_2) = (8-5)$$

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{16}{25} + \frac{1}{4}$$

$$\Rightarrow \frac{\bar{X}_1 - \bar{X}_2 - (8-5)}{\sqrt{16/25 + 1/4}} \sim N(0,1)$$

10.8

Betrakta kostnaderna för att reparera fiberoptik som gått sönder när det installerades respektive när systemet användes.

vid installation	vid användning
$n_1 = 21$	$n_2 = 25$
$\bar{X}_1 = 65$	$\bar{X}_2 = 120$
$S_1^2 = 25$	$S_2^2 = 100$

Man tror att varianserna för kostnaderna för reparation vid fel vid användning är större.

a) Ställ upp hypoteser för att testa detta.

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_2^2 > \sigma_1^2$

b) Använd detta för att testa vid alpha = 0.1

Sats 10.2.1: $S_2^2/S_1^2 \sim F(n_2-1, n_1-1)$ (under H_0)

Vi har $n_1=21$, $n_2=25$

Förkastar H_0 om S_2^2/S_1^2: $P(F \geq S_2^2/S_1^2) < 0.1$

$\frac{S_2^2}{S_1^2} > F_{0} \Rightarrow$ förkasta H_0

Tabell IX: $F_{0} = 1.767$

$\frac{S_2^2}{S_1^2} = \frac{100}{25} = 4 \Rightarrow$ förkastar H_0 vid 10% signifikans nivå.

10.10

Man har registrerat baspriserna på två olika ställen under en period ($/gallon).

South Carolina				
1.46	1.47	1.42	1.51	1.55
1.52	1.48	1.47	1.53	1.50

Michigan				
1.69	1.79	1.72	1.76	1.80
1.59	1.89	1.72	1.63	1.55
1.91	1.71			

Använd detta för att testa om variansen är lika.
Vad är p-värdet och slutsatsen?

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

$n_1 = 10$
$n_2 = 12$ $\Rightarrow$ $S_2^2 / S_1^2 \sim F(11, 9)$

$S_1^2 = 0.0015$

$S_2^2 = 0.012$

$\frac{S_2^2}{S_1^2} = 8.1$

tabell IV

$P(F_{11,9} > 8.1) < P(F_{11,9} > 3.102) = 0.05$

$\Rightarrow$ p-värde $< 0.1 \Rightarrow$ förkasta H_0

10.14
Jämför två laserskannare. Följande data fås för #streckkoder per sekund.

Ny	Gammal
$n_1 = 61$	$n_2 = 61$
$\bar{x}_1 = 40$	$\bar{x}_2 = 29$
$s_1^2 = 24.9$	$s_2^2 = 22.7$

a) Testa

$H_0: \sigma_1^2 = \sigma_2^2$ vid nivå $\alpha = 0.2$
$H_1: \sigma_1^2 \neq \sigma_2^2$

Förkastar H_0 om $s_1^2 / s_2^2 > f_{0,1}$ för $F(60,60)$

$s_1^2/s_2^2 = 1,097$, $f_{0,1} = 1.395$ ← tabell IX

Ok! H_0 förkastas ej.

b) Hitta

s_p^2?

$$s_p^2 = \frac{(n_2-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{60 \cdot 24,9 + 60 \cdot 22.7}{61+61-2} =$$

$= 23.8$

c) Hitta 90% CI för mu1-mu2

90% CI för $\mu_1 - \mu_2$

Sats 10.3.1: CI för $\mu_1 - \mu_2$

Pooled variance

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2} \sqrt{S_p^2 (1/n_1 + 1/n_2)}, \quad t_{\alpha/2} \text{ från } T_{n_1+n_2+2}$$

$t_{\alpha/2} \approx 1.660$ $(n = 100$ - egentligen $120)$ ty $\sigma_1^2 = \sigma_2^2$ antages

$\bar{X}_1 - \bar{X}_2 = 11 \Rightarrow CI = 11 \pm 1.660 \sqrt{23.8 \cdot 2/61} =$

$= [9.53; 12.47]$

d) verkar den nya scannern läsa fler streckkoder?

Ja, $0 \notin CI$.

e) Antalet streckkoder är diskret. Hur kan vi använda T-test?

Central Limit Theorem?

10.16

Visa att $E[S_p^2]=\sigma^2$

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

<u>vet</u>

$$\frac{(n_1-1)S_1^2}{\sigma^2}\sim\chi^2_{n_1-1} \qquad \text{(Sats. 8.1.1)}$$

$$\frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2_{n_2-1}$$

om $X\sim\chi_n \Rightarrow E[X]=n$ (kapitel 4)

$$E[S_p^2]=E\left[\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right]=$$

$$=\frac{1}{n_1+n_2-2}\left(E\left((n_1-1)S_1^2\right)+E\left((n_2-1)S_2^2\right)\right)$$

vet bara nått om $\frac{(n_1-1)S_1^2}{\sigma^2}$

men $(n_1-1)S_1^2=\sigma^2\cdot\frac{(n_1-1)S_1^2}{\sigma^2}$

$$E[S_p^2]=\frac{1}{n_1+n_2-2}\left(\sigma^2E\left[\frac{(n_1-1)S_1^2}{\sigma^2}\right]+\sigma^2E\left[\frac{(n_2-1)S_2^2}{\sigma^2}\right]\right)=$$

$$=\frac{1}{n_1+n_2-2}\left((n_1-1)\sigma^2+(n_2-1)\sigma^2\right)=\frac{n_1+n_2-2}{n_1+n_2-2}\sigma^2=\underline{\underline{\sigma^2}}$$

VSV.

10.24

En ny metod för att förvätska **kol testas. Följande data observeras från ny och gammal metod.**

ny - 1	gammal - 2
16.4 12.8	11.1 10.5 ....
17.7 ...	⋮ ...
⋮	

a) Ställ upp hypoteser för att testa om den nya metoden är bättre vid alpha=0.01

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

b) Testa hypotesen vid nivå alpha=0.01.

Lika varianser? Testa på nivå $\alpha = 0.1$:

$n_1 = 13 \quad s_1^2 = 5.89 \qquad \dfrac{s_1^2}{s_2^2} = 1.74, \quad f_{0.1;12,12} = 2.687$ tabell IX

$n_2 = 13, \quad s_2^2 = 3.39$

$\Rightarrow$ kan poola med gott samvete.

$\bar{x}_1 = 15.52$

$\bar{x}_2 = 12.85$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 4.638$$

Test statistiska: $T = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} = 3.1599$

tabell VI

kritiskt värde: $t_{n_1+n_2-2,\alpha} = t_{24;0.01} = 2.492$

$T > t_{24;0.01} \Rightarrow$ p-värde $< \alpha = 0.01$

Förkastar Ho och accepterar istället H_1, den nya metoden ger mer rekomenderar att byta.

10.27

CI för $\mu_1 - \mu_2$, olika varianser.

Ett $100(1-\alpha)\%$ CI för $\mu_1 - \mu_2$ när $\sigma_1 \neq \sigma_2$ ges av.

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{s_1^2/n_1 + s_2^2/n_2}$$

Finns det belägg för att bensinpriserna var högre i Michigan än i SC?

Michigan: $\bar{x}_1 = 1.73$, $s_1^2 = 0.012$, $n_1 = 12$

SC: $\bar{x}_2 = 1.49$, $s_2^2 = 0.0015$, $n_2 = 10$

$t_{\alpha/2}$? Fran $T[\gamma]$-fördelning

$$\gamma = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 14.12 \Rightarrow T_{14} \text{ fördelning}$$

$t_{\alpha/2} = 2.145$ — tabell VI

$\Rightarrow CI = [0.1664; 0.3116]$ — 95%-intervall

Ja det finns belägg, med 95% sannolikhet ligger det värdet mellan värdena i CI.

10.28

En tillverkare tillverkar något med komponenter som behöver tåla tryck. Komponenter från två leverantörer jämförs:

I	II
$n_1 = 10$	$n_2 = 10$
$\bar{x}_1 = 1350$	$\bar{x}_2 = 1338$
$s_1^2 = 100$	$s_2^2 = 29$

a) Vi vill ha ett 95% CI för mu1-mu2. Kan vi poola?

tabell IX

$S_1^2/S_2^2 \sim F(9,9)$, $f_{0,1} = 2.44$ $(\alpha = 0.2)$

$\frac{S_1^2}{S_2^2} = 3.45 > f_{0,1} \Rightarrow$ Antar olika varianser. kan inte poola.

b) Hitta ett 95% CI för mu1-mu2

95% CI $\mu_1 - \mu_2$.

$CI = \bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}\sqrt{S_1^2/n_1 + S_2^2/n_2}$

$t_{\alpha/2} = 2.262$ — tabell VI

$CI = 12 \pm 2.262\sqrt{\frac{100}{10} + \frac{29}{10}} = [4.16 ; 19.84]$

c) Finns det bevis för att komponenter från lverantör I i genomsnitt tål tryck?

Ja eftersom $0 \notin CI$

10.34

En ny metod för att halten av plutonium baserat på alpha-strålning testas. Man tror att den ger högre värde. En undersökning görs: Blanda lösning, dela i två och mät med respektive metod.

Uppmätt:

Sample	ny	Gammal	Skillnad
1	3,78	3.35	0.43
2	3.58	3,6	-0.02
3	3.77	3.41	0.36
4	⋮	⋮	⋮
5			
6			
7			
8			
9			
n = 10	3,53	3.61	-0.08

$\bar{D} = 0.151$

$S_d^2 = 0.0281$

Testa hypotesen $H_0: \mu_1 = \mu_2$ mot

$H_1: \mu_1 > \mu_2$

Dra en slutsats utifrån p-värdet.

Test statistika: $\frac{\bar{D}}{s_d/\sqrt{n}} \sim T_{n-1}$, $n = 10$

$\frac{\bar{D}}{s_d/\sqrt{n}} = 2.8475$, $P\left(T > \frac{\bar{D}}{s_d/\sqrt{n}}\right) = 0.0096 < 0.01$

skulle förkasta H_0 vid signifikansnivå 0.01

10.43

Ett företag har två leverantörer att välja på. Man misstänker att X tar mer betalt än Y för liknande produkter. Stöder data detta vid alpha = 0.05?

X	Y	|X-Y|	Rank	Signed Rank
6000	5900	100	7.5	7.5
575	580	5	2	-2
15000	15000	0	1	-1
150000	145000	5000	10	10
76000	75000	1000	9	9
5650	5600	50	6	6
10000	9975	25	5	5
850	870	20	4	-4
900	890	10	3	3
3000	2900	100	7.5	7.5

$H_0: M_X = M_Y$ $\quad$ w_- test statistika

$H_1: M_X > M_Y$

$\frac{10\cdot(10+10+1)}{2}$

$W_- = 1+2+4 = 7$

Kritiskt värde: Tabell $\Rightarrow$ $= 105 - 83.13 = 11$

$W_- < 11$ Så förkastar H_0.

Skulle sign test ge samma resultat?

Sign test: Räkna # negativa differanser,

Q_- där vi observerat $q = 3$

Under H_0 $Q_- \sim Bin(10, 0.5)$

p-värde $= P(Q_- \leq q)$

tabell I

Vi har $q = 3$ $\quad P(Q_- \leq 3) \stackrel{\downarrow}{=} 0.1719 > 0.05$

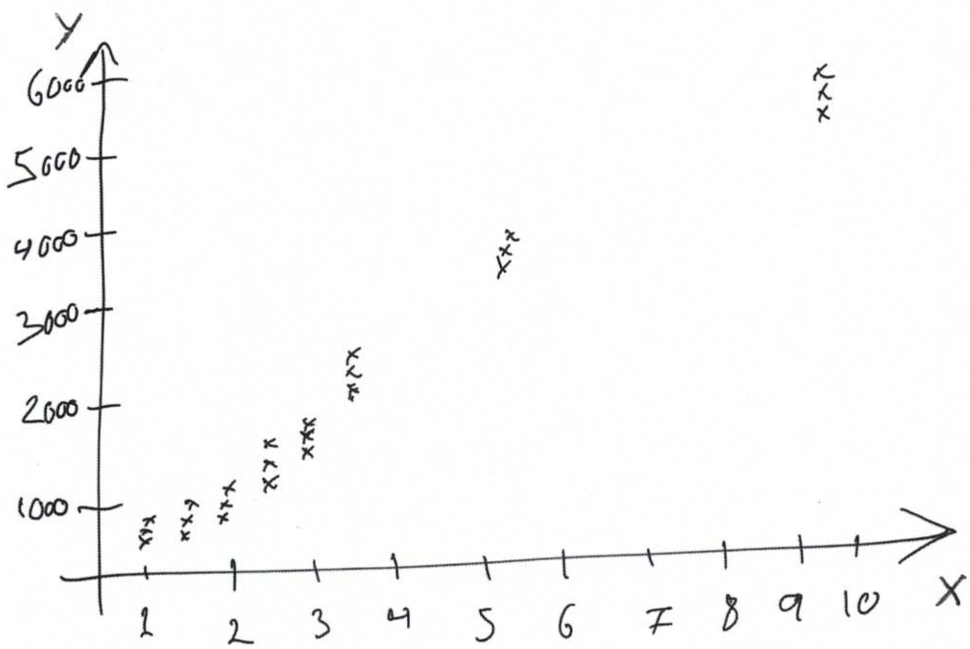
Kan nu inte förkasta H_0.

Skillnad är att vi här inte tar hänsyn till storlek.

11.10
En power brush används för att ta bort ojämnheter på metall. En borste testas vid olika rpm.

X (1000 rpm)	Y (täkt yta)
1	525, 520, 527
1.5	785, 780, 790
1.75	915, 900, 922
2.5	1300, 1295, 1310
3	1575, 1565, 1582
4	2100, 2110, 2090
6	3125, 3120, 3133
10	5250, 5256, 5245

a) Skissa en skatter plot.



b) skatta regressions kurvan.

Kurvan ges av $y = b_0 + b_1 x$

där $b_0 = \bar{y} - b_1 \bar{x}$

$$b_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

$\sum x_i = 89.25$, $\sum x_i^2 = 520.69$, $\sum y_i = 46720$

$n = 24$ $\quad \sum x_i y_i = 272801$

$\Rightarrow b_1 = 524.71, \quad b_0 = -4.595$

c) skatta hur stor yta som täcks om vi kör med 3450 rpm.

Använder regressions modellen från b)

$\hat{y} = b_0 + b_1 x = -4.595 + 524.71 \cdot 3.450 = 1806$

11.20

Hitta 95% CI för beta0. Data utifrån uppgift 7.

95% CI för β_0 $\quad n = 8$

$\bar{x} = 58.05$

$\sum x_i^2 = 32090$

$\bar{y} = 5.775$

$\sum x_i y_i = 3173.2$

$b_1 = 0.09573$

$b_0 = 0.2177$

$$CI = \beta_0 \pm t_{\alpha/2} \frac{s\sqrt{\sum x^2}}{\sqrt{n\, S_{xx}}}$$

$$s^2 = \frac{SSE}{n-2}, \quad SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

tabell VI

$$SSE = 1.505, \quad S_{xx} = 5131.54, \quad t_{0.025;6} = 2.447$$

$$\Rightarrow CI = 0.2177 \pm 1.0835 = [-0,8658; 1,3013]$$

11.22

a) om X=50 skatta ett värde för Y.

$$\hat{y} = 0.2177 + 0.09573 \cdot 50 = 5.00$$

b) ta fram ett 95% CI för y givet x=50 och tolka svaret.

$$\hat{Y}|x \pm t_{\alpha/2}\, S\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

$t_{\alpha/2} = 2.447$

$x = 50$

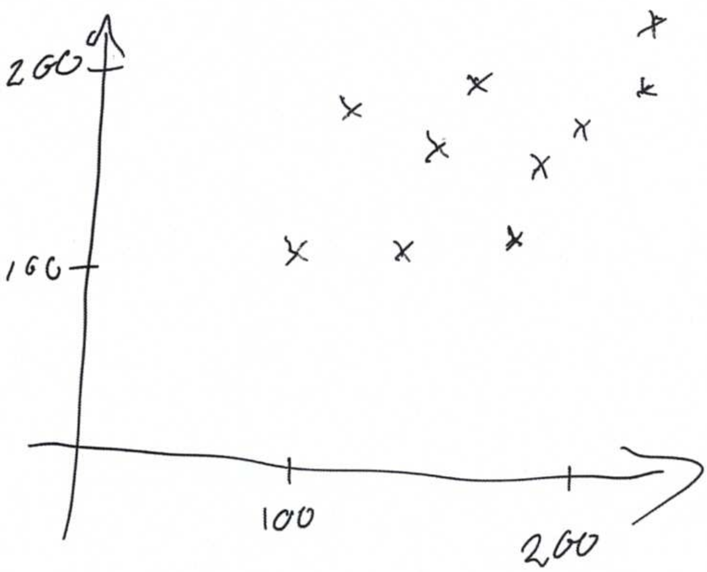
$$CI = 5.00 \pm 1.3071 = [3.69; 6.31]$$

11.47

Bekämpningsmedel hittas i mat. Man gör en studie där kyckling får äta madaoxon. De fick också en lever enzyms indexerade för att testa om leverns avgiftnings process påverkades. Följande data observerades som procentandel av normal bekämpningsmedelsavgiftning (y) och procentandel av normal enzymnivå (x)

Enzymnivå (x)	Avgiftningsnivå (y)
95	108
110	126
118	102
124	121
145	118
140	155
185	158
190	178
205	159
222	184

a) Plotta datan



ser ut att vara möjligt att göra regrision.

b) Skatta rho, korrelationen mellan X och Y.

ρ

$$\hat{\rho} = r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - \left(\sum x\right)^2\right)\left(n\sum y^2 - \left(\sum y\right)^2\right)}}$$

$n = 10 \quad \sum x = 1534 \quad \sum y = 1409 \quad \sum xy = 226463$

$\sum x^2 = 252684 \quad \sum y^2 = 206319$

$\Rightarrow \quad \underline{\underline{r = 0.8874}}$

11.48

Tensta nollhypotesen att X och Y inte är korrelerade dvs H0: rho=0, vid alpha = 0.1.

$H_0: \rho = 0$

$H_1: \rho \neq 0 \qquad \alpha = 0.1$

Test statistika $\frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim T_{n-2}$

Vårt värde

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 5.4438$$

$$P(|T| > 5.4438) = P(T < -5.44) + P(T > 5.44) =$$

$$= 2 \cdot P(T > 5.44) \leq 0.001$$

Förkastar alltså H_0 och kan dra slutsatsen att enzymhalten påverkar avgiftningen.