

Spørsmål

17 mars 2004

1) Kan du gi en forklaring på hva som menes med "nullhypotese" og at denne forkastes? Vil bare ha det oppklart.

- Nullhypotese er ein påstand vi gjer om ein parameter for å kunne teste om det er ein rimeleg korrekt påstand
- Hvis nullhypotesa leier til ein urimeleg konklusjon (t.d. stor t-verdi) forkastar vi den som urimeleg
- Uttrykket kan kanskje ha samanheng med at vi vanlegvis tar utgangspunkt i at parameteren er 0, men generelt vil det i dag bli brukt om
- "Et utsagn hvis feilaktige benektelse vi fortrinnsvis ønsker å unngå" (Sverdrup 1964:175 "Lov og tilfeldighet I")
- Synspunktet er at det er best å gjere feil på rett side, t.d. i kvalitetskontroll av fly, bilar, klatretau, ...

2) Når man skal definere relasjonen mellom variablene i en modell, skal "e" være inkludert også for at det skal være korrekt?

Svaret er ja! Men legg merke til:

- Observert $Y_i = \text{Predikert } Y_i + e_i$
(Predikert $Y_i = \text{Forventa verdi av } Y_i = E[Y_i]$)

Modellen kan omtalast på to måtar

- $E[Y_i] = f(x_i)$ da er residualen underforstått eller

- $Y_i = f(x_i) + e_i = E[Y_i] + e_i$ da er residualen inkludert eksplisitt

Og hugs skiljet mellom populasjon og utval

3) Litt forvirring ang multikollinearitet og autokorrelasjon, og hvordan vi kan finne ut om det er det.

Substansielt:

- Autokorrelasjon dreiar seg om tilhøvet mellom verdier på ulike case på same variabel (t.d. prosent yrkesaktive fiskarar i kvar kommune). Geografisk nærliggjande case har ofte verdier som ligg nærmare kvarandre enn to tilfeldig valde kommunar
- Multikollinearitet dreiar seg om korrelasjon mellom variable (t.d. "tillit til Stortinget" og "tillit til politikarar"). Variablar som måler om lag same fenomenet **skal** korrelere med kvarandre og er til ein viss grad gjensidig overflødige
- Teknisk ?

4) Er Gauss Markov kravene oppfylt selv om det ikke er normalfordelte restledd? Kan f- og t-test brukes selv om restleddene ikke er normalfordelte?

- Gauss-Markov krava kan godt vere oppfylt sjølv om vi ikkje har normalfordelte restledd, men dei treng ikkje vere det
- F-test og t-test kan nyttast **berre** dersom restleddet er normalfordelt

5) Oppgave 1, Høst 2003, Hypotese 2 - "female" er IKKE signifikant. Betyr det at H2 forkastes, eller at man ikke kan si noe om hypotesen?

- I modell 2 og 3 er "female" ikkje signifikant. Vi kan da ikkje forkaste da hypotesen om at "female" ikkje har effekt. Det er med andre ord ikkje grunnlag for å påstå skilnad mellom menn og kvinner. Når føresetnadene er oppfylt veit vi at sjansen for at dette er feil konklusjon er mindre enn 5% (signifikansnivået er 0.05)
- Er det verre å påstå forskjell dersom det ikkje er det enn å påstå at det ikkje er forskjell når det faktisk er det? (jfr første spørsmålet)

6) Skal man KUN se på signifikansen når man vurderer riktigheten av hypoteser?

- Ved sida av signifikansnivå må ein alltid vurdere tolkinga substansielt. Dersom det vi finn strir mot sunn fornuft, eller det andre forskarar har funne bør vi sjekke etter kva vi har gjort ein gong til ...
- Det gjeld sjølvsagt både for signifikante og usignifikante variablar

7) Ihht forelesningsnotater og Hamilton er forutsetning 1 for at en gitt modell er korrekt at modellen må være "lineær i parameter". Hva betyr det? Er det et problem med feks Age square?

Legg merke til at kravet ikkje går på om modellen er korrekt eller ikkje. Kravet går på om vi skal kunne nytte OLS metoden for å estimere parametrane

- **Parametrane** i lineære modellar er β_k eller " b_k "-ar som vi estimerer
- "Lineær i parametrane" tyder at vi finn ein og berre ein " b_k " mellom kvart "+" i modellen
- Sidan det står berre ein " b_k " framfor "AgeSquare" er modellen lineær i parameteren

8) Hva sier tabellen som heter "casewise diagnostics"?
Hvordan og når skal man bruke den?

Casewise Diagnostics^a

Case Number	Std. Residual	SYSTRUST Trust in system, mean of b7-b10	Predicted Value	Residual
11883	-3.046	.00	5.2601	-5.2601
28014	-3.125	.75	6.1475	-5.3975
28963	-3.303	.00	5.7047	-5.7047
29216	-3.190	.00	5.5097	-5.5097
29438	-3.029	.75	5.9811	-5.2311
29651	3.280	9.25	3.5858	5.6642
29781	3.657	10.00	3.6848	6.3152
30854	3.313	9.50	3.7784	5.7216

a. Dependent Variable: SYSTRUST Trust in system, mean of b7-b10

8) "casewise diagnostics" tabellen

- Bestiller vi "casewise diagnostics" får vi ein tabell med variabelverdier for dei case som møter kriteriet som er gitt ("outliers outside n standard deviations").
- Default-verdien av n er 3.
- Variablane som er skrive ut er y (avhengig variabel), predikert y, residual, og standardisert residual (seleksjonsvariabel)
- Tabellen vil seie oss om utliggjara er eit stort eller lite problem

9) Hva er en P-P Plot? Finner ikke i Hamilton (om den ikke har annet navn).

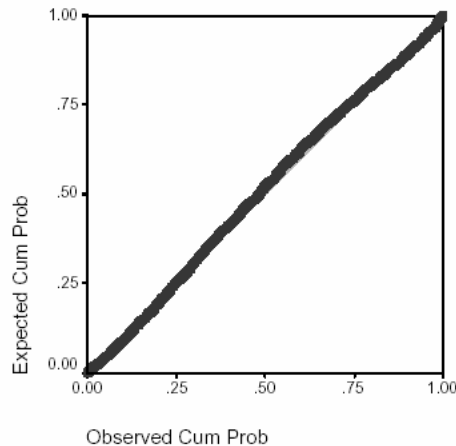
- I regresjonsprosedyren er det noko som heiter "Normal P-P Plot" og i følgje hjelpemenyen vil dette
- "Displays a normal probability plot of the standardized residuals. Used to check for normality. If the variable is normally distributed, the plotted points form a straight diagonal line."
- Dette er **ikkje** det same som det Hamilton kallar "Quantil-Normal Plots" (sjå side 15 i Hamilton) men har same formål: å teste om ei observert fordeling er lik normalfordelinga

Normal P-P Plot i regresjon:
normal probability plots comparing the distribution of standardized residuals to a normal distribution

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: Trust in system, mean of b7-b10

Forventa ved
Normalfordeling



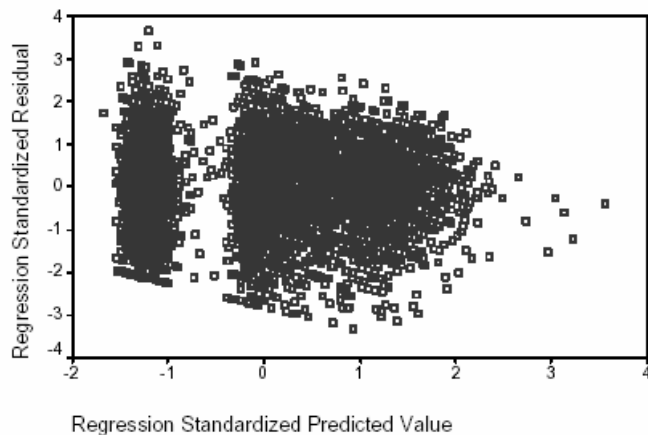
P-P og Q-Q plott i "Graphs" menyen

- P-P Plot
- Plots a variable's **cumulative proportions** against the cumulative proportions of any of a number of test distributions. Probability plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line.
- Q-Q Plot
- Plots the **quantiles of a variable's distribution** against the quantiles of any of a number of test distributions. Probability plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line.

10) Scatterplotten på side 8 i Høst 2003 eksamen - hva sier denne? Den har jo så merkelig form.

Scatterplot

Dependent Variable: Trust in system, mean of



Merkjeleg spreingsdiagram?

- Det finst i datamaterialet grupper som vi ventar skal ha systematisk lågare predikert verdi av "Trust in system" (tre land: Polen, Norge, Storbritannia)
- Avhengig variabel har avgrensingar i variasjonsområdet sitt. Den kan t.d. ikkje bli mindre enn 0 eller større enn 10. Dette fører til systematikk i korleis residualane varierer, særleg der observert verdi ligg nær yttergrensene. Vi får "rette linjer" i utkanten av variasjonsområdet

11) Hva vil det si at regresjonskoeffisienten skal være signifikant forskjellig fra null? Hvor kontrollerer man det?

- I SPSS utskrifta i tabellen "Coefficients" finn vi ei kolonne merka "t" (Students's t-observator) og ei merka "Sig." (sannsynet for så stor eller større |t|)
- På neste slide står det Sig. =0.268 for variabelen "Female". Dette les vi slik:
- Dersom nullhypotesa (ingen effekt av "Female") er rett er det eit sannsyn på 0.268 for å finne ein t-verdi som er større eller lik 1.107 eller mindre eller lik -1.107
- Sidan verdien av observatoren synest rimeleg under gitte føresetnader trur vi at føresetnadene er rimelege

Haust 2003 1 e)

Model	Dependent Variable: SYSTRUST Trust in system, mean of b7-b10 a.	Unstandardized Coefficients		Standardized Coefficient	t	Sig.	Collinearity Statistics	
		B	St.Err				Beta	Tolerance
	(Constant)	4.246	.199		21.313	.000		
	FEMALE	-.049	.045	-.013	-1.107	.268	.992	1.008
	AGE	-.042	.007	-.389	-5.932	.000	.029	33.909
	AGE2 Age squared	.000	.000	.408	6.274	.000	.030	33.264
	MARRIED Married (1), not married (0)	.118	.051	.030	2.334	.020	.767	1.304
	EDUYRS Years of full-time education completed	.020	.012	.037	1.743	.081	.287	3.487
	GB Dummy for Great Britain	.751	.202	.180	3.725	.000	.054	18.391
	NO Dummy for Norway	.856	.201	.205	4.269	.000	.055	18.199
	EDUK ED*GB interaction	.034	.016	.110	2.136	.033	.048	20.814
	EDNO ED*no interaction	.102	.016	.340	6.496	.000	.046	21.584

12) Hva betyr "Stokastisk variasjon"?

- Stokastisk refererer til sannsyn
- I uttrykket "stokastisk variasjon" tenker ein seg variasjon t.d. i form av avvik frå eit gjennomsnitt som det kan knytast visse sannsyn til: t.d. til større avvik til mindre sannsyn
- I uttrykket "stokastisk variabel" vil det kunne knytast sannsyn til ulike variabelverdier.