

# SOS3003

## Eksamensoppgåver

Gjennomgang våren 2004

Erling Berge

## Gjennomgang av

- Oppgåve 1 gitt hausten 2003

# Haust 2003 Oppgave 1

Den avhengige variabelen i regresjonsanalysen er en skala (indeks) for tillit til samfunnsinstitusjoner. Se ellers dokumentasjonen i tabellvedlegget.

- a) Hvilken av de tre modellen er best?
- b) Definer modellen for populasjonen (uten forutsetninger). Ta utgangspunkt i den beste modellen.
- c) Er det forskjeller mellom landene når det gjelder tillit til samfunnsinstitusjoner?
- d) Lag et betinget effektdiagram som viser effekten av utdanning for hvert av de tre landene.
- e) Vurder følgende hypoteser:
  - H1: Alder har en lineær effekt på tillit til samfunnsinstitusjoner.
  - H2: Kvinner har mindre tillit til samfunnsinstitusjoner enn menn.
  - H3: Utdanning har ingen effekt på tillit til samfunnsinstitusjoner i Polen.
- f) Vurder om forutsetningene til modell 3 er tilfredsstillt.

## Haust 2003 1 a) Hvilken av de tre modellene er best?

### ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1112.506	5	222.501	60.251	.000
	Residual	22379.014	6060	3.693		
	Total	23491.520	6065			
2	Regression	5296.746	7	756.678	251.938	.000
	Residual	18194.774	6058	3.003		
	Total	23491.520	6065			
3	Regression	5428.895	9	603.211	202.243	.000
	Residual	18062.625	6056	2.983		
	Total	23491.520	6065			

Tabell frå vedlegget

Haust 2003 1 a)  
Test av to modeller mot kvarandre

$$F_{n-K}^H = \frac{(RSS\{K-H\} - RSS\{K\})/H}{(RSS\{K\})/(n-K)}$$

Haust 2003 1a) Tre modeller

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1112.506	5	222.501	60.251	.000a
	Residual	22379.014	6060	3.693		
	Total	23491.520	6065			
2	Regression	5296.746	7	756.678	251.938	.000b
	Residual	18194.774	6058	3.003		
	Total	23491.520	6065			
3	Regression	5428.895	9	603.211	202.243	.000c
	Residual	18062.625	6056	2.983		
	Total	23491.520	6065			

Haust 2003 1 a)  
Kva for ein av dei tre modellane er best?

Er modell 2 bedre enn modell 1?			Er modell 3 bedre enn modell 2?	
K	8		K	10
H	2		H	2
RSS(K-H):	22379,01		RSS(K-H):	18194,77
RSS(K):	18194,77		RSS(K):	18062,63
n	6066		n	6066
df1=H	2		df1=H	2
df2=n-K	6058		df2=n-K	6056
<b>F</b>	<b>696,577</b>		<b>F</b>	<b>22,153</b>

Haust 2003 1 b) Modellen for  
populasjonen (uten forutsetninger)

- Definer modellen for populasjonen (uten forutsetninger). Ta utgangspunkt i den beste modellen.

Når vi skal formulere ein modell må vi

1. definere elementa som inngår i modellen (variablar og datamaterialet)
2. definere relasjonane mellom elementa (regresjonslikninga), og
3. presisere kva føresetnader som ein må gjere for å bruke modellen (spesifikasjon av modell, Gauss-Markov krav, normalfordelte feilledd)

## Haust 2003 1 b) Elementa i modellen

- Populasjonen vi granskar er befolkningane i dei tre landa Storbritannia (GB), Norge (NO) og Polen (PO)
- Det skal lagast ein modell av den avhengige variabelen  $Y$  = tillit, ein skala (indeks) for tillit til samfunnsinstitusjonar i landet til respondenten (gjennomsnittet av variablane b7-b10 i datamaterialet)
- Den beste modellen er i følgje 1a) modell 3
- Det vil bli argumentert med at følgjande uavhengige variablar verkar inn på korleis ein svarar på spørsmålet:

## Haust 2003 1 b) Elementa i modellen

Variabel symbol	Variabel: namn og tolking, "i" identifiserer eintydig ein respondent i populasjonen
$X_{1i}$	FEMALE = 1 dersom respondenten i er kvinne, 0 elles
$X_{2i}$	AGE = alder i år for respondenten i
$X_{3i}$	AGE2 = AGE kvadrert
$X_{4i}$	MARRIED = 1 dersom respondenten i er gift, 0 elles
$X_{5i}$	EDUYRS = talet av år med full tids utdanning for respondenten i
$X_{6i} - X_{7i}$	Dummykoding av CNTRY3 gir 2 variablar i modellen CNTRY3 = 1 dersom respondenten i er frå Great Britain gir GB CNTRY3 = 2 dersom respondenten i er frå Norge gir NO, CNTRY3 = 3 dersom respondenten i er frå Polen gir referansekategori
$X_{8i} - X_{9i}$	Interaksjon mellom EDUYRS og CNTRY3 gir 2 variablar i modellen

## Haust 2003 1 b) Modellen (1)

- I populasjonen føreset vi at det er eit lineært eller kurvelineært samband mellom den avhengige variabelen, Y, og dei uavhengige X-variablane. Dette tyder i vårt høve at når vi lar i gå over heile populasjonen vil
- $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i} + \varepsilon_i$$

## Haust 2003 1 b) Modellen (2)

- Modellen kan og skrivast
- $$Y_i = \beta_0 + \beta_1 \text{FEMALE}_i + \beta_2 \text{AGE}_i + \beta_3 \text{AGE2}_i + \beta_4 \text{MARRIED}_i + \beta_5 \text{EDUYRS}_i + \beta_6 \text{GB}_i + \beta_7 \text{NO}_i + \beta_8 (\text{EDUYRS}_i * \text{GB}_i) + \beta_9 (\text{EDUYRS}_i * \text{NO}_i) + \varepsilon_i$$

## Haust 2003 1 b) (ekstrastoff) føresetnadane (1)

Vi kan estimere dei ukjente parametrane i denne modellen dersom vi har observasjonar for eit reint tilfeldig utval frå populasjonen og vi kan gjere følgjande føresetnader:

- I. Modellen er korrekt, dvs.:
  - alle relevante variablar er med
  - ingen irrelevante er med
  - modellen er lineær i parametrane

## Haust 2003 1 b) (ekstrastoff) føresetnadane (2)

- II. Gauss-Markov krava for «Best Linear Unbiased Estimates» (BLUE) er oppfylt, dvs.:
  - Faste x-verdiar (dvs. vi kan i prinsippet trekke nye utval med same x-verdiar men der vi får ulike y-verdi på grunn av den stokastiske komponenten i feilleddet).
  - Feilledda har forventning 0 for alle  $i$ , dvs:  $E(e_i) = 0$  for alle  $i$ .
  - Feilledda har konstant varians (homoskedastisitet) dvs:  $\text{var}(e_i) = \sigma^2$  for alle  $i$ .
  - Feilledda er ukorrelerte med kvarandre (det er ikkje autokorrelasjon) dvs:  $\text{cov}(e_i, e_j) = 0$  for alle  $i \neq j$ .
- III. Normalfordeling av feilleddet:
  - Feilledda er normalfordelte med same varians for alle case, dvs:  $e_i \sim N(0, s^2)$  for alle  $i$ .

## Haust 2003 1 c) Er det forskjeller mellom landene når det gjelder tillit til samfunnsinstitusjoner?

- Kva slags test kan gje svar på dette spørsmålet
- Kva for ein tabell i oppgåva gir oss opplysningar om skilnader mellom landa?

H2003 1 c)	Dependent Variable: SYSTRUST Trust in system, mean of b7-b10 a.	Unstandardized Coefficients		Standardized Coefficient			Collinearity Statistics	
		B	St.Err	Beta	t	Sig.	Tolerance	VIF
Model								
1	(Constant)	3.406	.183		18.614	.000		
	FEMALE	-.105	.049	-.027	-2.120	.034	.996	1.004
	AGE	-.012	.008	-.114	-1.570	.116	.030	33.443
	AGE2 Age squared	.000	.000	.207	2.868	.004	.030	33.037
	MARRIED Married (1), not married (0)	-.146	.056	-.037	-2.623	.009	.790	1.266
	EDUYRS Years of full-time education completed	.125	.007	.225	16.873	.000	.884	1.132
2	(Constant)	3.667	.165		22.202	.000		
	FEMALE	-.044	.045	-.011	-.993	.321	.993	1.008
	AGE	-.041	.007	-.377	-5.722	.000	.030	33.880
	AGE2 Age squared	.000	.000	.398	6.097	.000	.030	33.246
	MARRIED Married (1), not married (0)	.105	.051	.027	2.069	.039	.769	1.301
	EDUYRS Years of full-time education completed	.068	.007	.123	9.917	.000	.834	1.199
	GB Dummy for Great Britain	1.125	.057	.269	19.843	.000	.693	1.442
	NO Dummy for Norway	2.117	.057	.507	37.325	.000	.692	1.445
3	(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



**Haust 2003 1 d)** Lag et betinget effekt diagram som viser effekten av **utdanning** for hvert av de tre **landene**

- Vi skal studere sammenhengen mellom utdanning og tillit til samfunnsinstitusjonar
- Vi har estimert ein samanheng
- $Y_i = 4.246 - 0.049FEMALE_i - 0.042AGE_i + 0.0000001AGE2_i + 0.118MARRIED_i + 0.020EDUYRS_i + 0.751GB_i + 0.856NO_i + 0.034(EDUYRS_i * GB_i) + 0.102(EDUYRS_i * NO_i) + e_i$

**Haust 2003 1 d)** Betinga effektplott kan vise effekten av **utdanning** for hvert av de tre **landene**

- Betinging: kva andre variablar verkar ved sida av utdanning og land?
  - kjønn, alder, ekteskapeleg status
- Kva verdiar set vi inn for å få til effektplottet
- Vi treng 3 plott eitt for GB=1, eitt for NO=1 og eitt for GB = NO = 0

## Haust 2003 1 d) Betinga effektplott

- Om vi ryddar litt i dei ulike variablane her får vi
- $Y_i = 4.246 - 0.049FEMALE_i - 0.042AGE_i + 0.0000001AGE2_i + 0.118MARRIED_i + 0.751GB_i + 0.856NO_i + [0.020 + 0.034(GB_i) + 0.102(NO_i)]EDUYRS_i + e_i$
- For å kunne plote ut ein samanheng må vi finne fram til høvelege verdiar på variablane kjønn, alder og ekteskapeleg status

## Haust 2003 1 d) variablane

### Gjennomsnitt

FEMALE	0,498187
AGE	45,60765
AGE2 Age squared	2406,482
MARRIED Married (1), not married (0)	0,52786
EDUYRS Years of full-time education completed	12,48698
GB Dummy for Great Britain	0,333333
NO Dummy for Norway	0,333663
EDUK ED*GB interaction	4,239202
EDNO ED*no interaction	4,419552

## Haust 2003 1 d) betinging effekt av utdanning i ulike land

Predikert verdi av SYSTRUST	B	Betinging ved å sette inn variabel verdi gjnsnitt	Alternativ betinging
(Constant)	4,246	--	--
FEMALE	-0,049	0,5	0
AGE	-0,042	45,6	50
AGE2 Age squared	0	2406,48	2500
MARRIED Married (1), not married (0)	0,118	0,527	0
EDUYRS Years of full-time education completed	0,02	0-20	0-20
GB Dummy for Great Britain	0,751	0-1	0-1
NO Dummy for Norway	0,856	0-1	0-1
EDUK ED*GB interaction	0,034		
EDNO ED*no interaction	0,102		

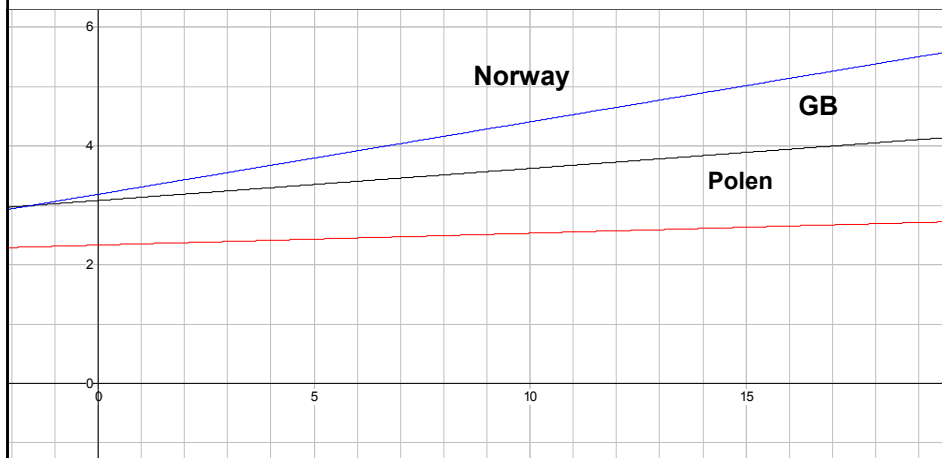
## Haust 2003 1 d) Betinga effektplott

- Vel vi å samanlikne ugifte menn på 50 år (nb: ikkje gjennomsnitt) vil likninga som skal plottast bli
- Predikert  $Y_i = 4.246 - 0.049 \cdot 0 - 0.042 \cdot 50 + 0.0000001 \cdot 50 \cdot 50 + 0.118 \cdot 0 + 0.751GB_i + 0.856NO_i + [0.020 + 0.034(GB_i) + 0.102(NO_i)]EDUYRS_i$
- Predikert  $Y_i = 2.146 + 0.751GB_i + 0.856NO_i + [0.020 + 0.034(GB_i) + 0.102(NO_i)]EDUYRS_i$   
der vi set in 1 eller 0 for å finne dei ulike landa

## Haust 2003 1 d) Betinga effektplott, utrekning

country	EDUYRS= 0	EDUYRS= 10	EDUYRS= 20
GB=1 (=> NO=0)	2.897	3.437	3.977
NO=1 (=> GB=0)	3.002	4.222	5.442
GB=NO=0 (=> Polen)	2.146	2.346	2.546

### Haust 2003 1 d) Effekten av utdanning for ugifte menn med gjennomsnittleg alder i Polen, Storbritannia og Norge



$$y = 4.246 - 0.049x + 0.0042x^2 + 0.0000001x^3 + 2406.4817 + 0.118x + 0.02x + 0.751 + 0.034x$$

$$y = 4.246 - 0.049x + 0.0042x^2 + 0.0000001x^3 + 2406.4817 + 0.118x + 0.02x + 0.856 + 0.102x$$

$$y = 4.246 - 0.049x + 0.0042x^2 + 0.0000001x^3 + 2406.4817 + 0.118x + 0.02x$$

## Haust 2003 1 e) Vurder følgende hypotesar:

- H1: Alder har en lineær effekt på tillit til samfunnsinstitusjoner.
- H2: Kvinner har mindre tillit til samfunnsinstitusjoner enn menn.
- H3: Utdanning har ingen effekt på tillit til samfunnsinstitusjoner i Polen.

## 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				Tolerance	VIF
	(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

## Haust 2003 1 f) Vurder om forutsetningene til modell 3 er tilfredsstillt.

### Kva er føresetnadene?

- I. Modellen er korrekt, dvs.:
  - alle relevante variablar er med
  - ingen irrelevante er med
  - modellen er lineær i parametranne
- II. Gauss-Markov krava
  - Faste x-verdiar
  - Feilledda har forventning 0 for alle i, dvs:  $E(\varepsilon_i) = 0$  for alle i .
  - Feilledda har konstant varians, dvs  $\text{var}(\varepsilon_i) = \sigma^2$  for alle i
  - Feilledda er ukorrelerte med kvarandre dvs:  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  for alle  $i \neq j$
- III. Normalfordeling av feilleddet:
  - Feilledda er normalfordelte med same varians for alle case, dvs:  $\varepsilon_i \sim N(0, \sigma^2)$  for alle i .

## Haust 2003 1 f) Vurder om forutsetningene til modell 3 er tilfredsstillt.

- Kva kan vi i praksis undersøkje?
  - I. Modellen er korrekt ?
    - Irrelevante variablar?
  - II. Gauss-Markov krava ?
    - Homoskedastisitet?
    - Autokorrelasjon?
  - III. Normalfordeling av feilleddet ?
  - Andre tekniske problem
    - Multikollinearitet
    - Case med stor påverknad

## Haust 2003 1 f) Irrelevante variablar? Multikollinearitet?

Dependent Variable: SYSTRUST Trust in system, mean of b7-b10 a.		Unstandardized Coefficients		Standardized Coefficient			Collinearity Statistics	
Model		B	St.Err	Beta	t	Sig.	Tolerance	VIF
3	(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

## Haust 2003 1 f) Case med påverknad?

Casewise Diagnostics<sup>a</sup>

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

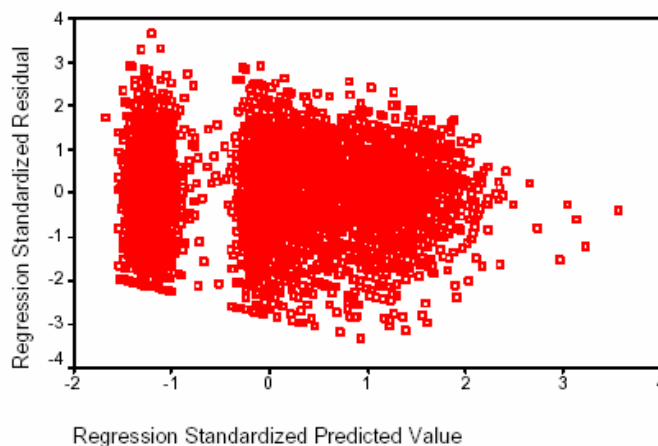
## Haust 2003 1 f) Ikkje Autokorrelasjon?

- Oppgåveteksten seier: ”I begge eksamenoppgåvene vert det nytta data frå European Social Survey (ESS). Føremålet med ESS er å kartlegge samanhengen mellom institusjonelle tilhøve i dei europeiske landa og borgarane sine holdningar, verdiar og oppfatningar. Den første intervjurunden vart utført hausten 2002 og data var gjevne fri for allmenn bruk hausten 2003. I alt har 22 land teke del i undersøkinga. I dei to eksamensoppgåvene er talet på land avgrensa til tre: Storbritannia, Noreg og Polen.”
- I surveygranskingar har vi ikkje grunnlag for å tru det skal finnast autokorrelasjon

## Haust 2003 1 f) Homoskedastisitet?

Scatterplot

Dependent Variable: Trust in system, mean of

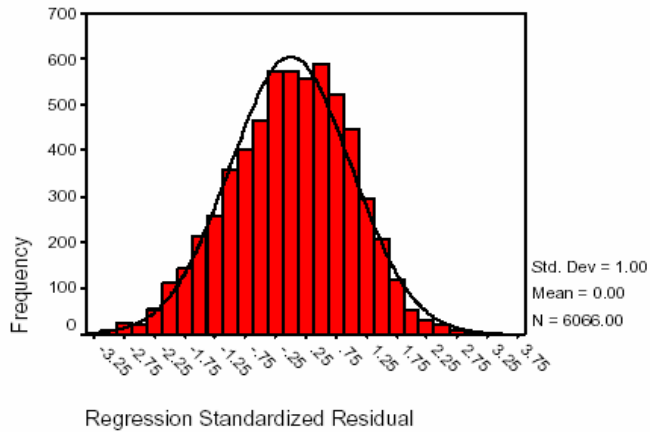




# Haust 2003 1 f) Normalfordelte residualar?

Histogram

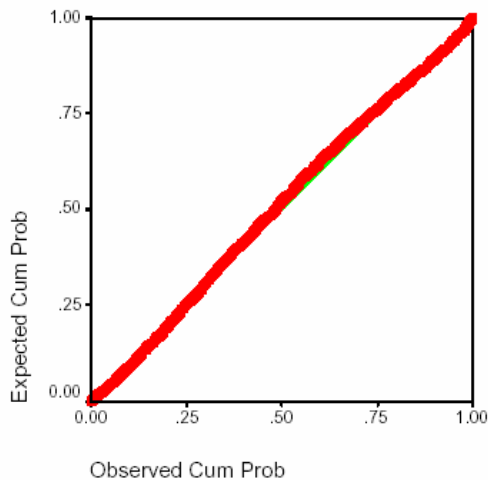
Dependent Variable: Trust in system, mean of



# Haust 2003 1 f) Normalfordelte residualar?

Normal P-P Plot of Regression Standardized Residual

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



## Haust 2003 1 f) Konklusjon

- Dei føresetnadene vi kan studere synest rimeleg godt oppfylt
  - Residualane er tilnærma normalfordelte med nokre få utliggjarar
  - Det er lite spor av heteroskedastisitet
  - Det er ikkje noko hypotese om autokorrelasjon
  - Det er ein irrelevant variabel
- Av andre tekniske problem noterer vi
  - Det er ein del multikollinearitet bygd inn i modellen
  - Det er vanskeleg å avgjere om utliggjarar har urimeleg stor påverknad