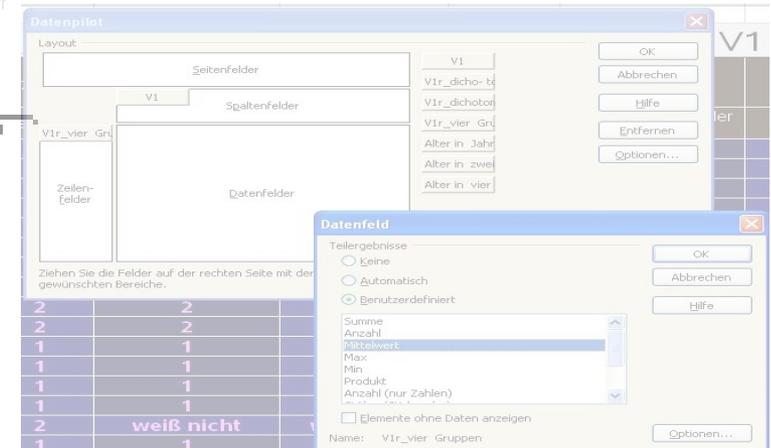
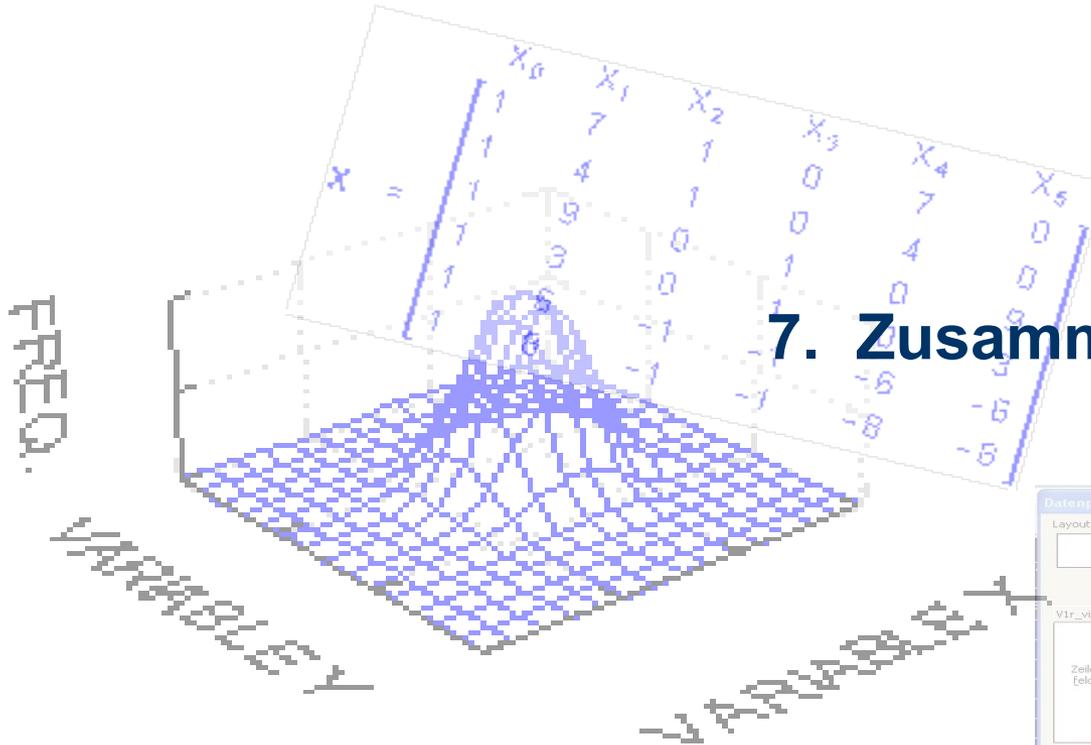


## Statistik IV

### 7. Zusammenhangsanalysen III und Probeklausur



# Terminplanung

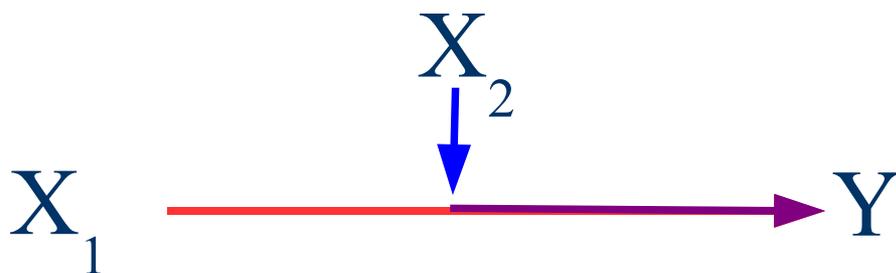
<i>Nr.</i>	<i>Termin</i>	<i>Inhalt</i>
1	14.04.09	<b>Einführung</b> <ul style="list-style-type: none"> <li>• Organisatorisches und Scheinvoraussetzungen</li> <li>• Statistik mit Softwareunterstützung?</li> </ul>
2	28.04.09	<b>Grundlagen Stata</b> <ul style="list-style-type: none"> <li>• Aufbau des Programms</li> <li>• Umgang mit Daten, Dokumentation</li> </ul> <b>Datenaufbereitung für statistische Auswertungen I</b>
3	12.05.09	<b>Datenaufbereitung für statistische Auswertungen II</b>
4	26.05.09	<b>Deskriptive Statistik und graphische Darstellung</b>
5	09.06.09	<b>Zusammenhangsanalyse I</b>
6	23.06.09	<b>Zusammenhangsanalyse II</b>
7	07.07.09	<b>Zusammenhangsanalyse III</b> <b>Probeklausur</b> <b>Besprechung der Probeklausur</b>
8	14.07.09	<b>Klausur (90 Minuten)</b>

# Übersicht

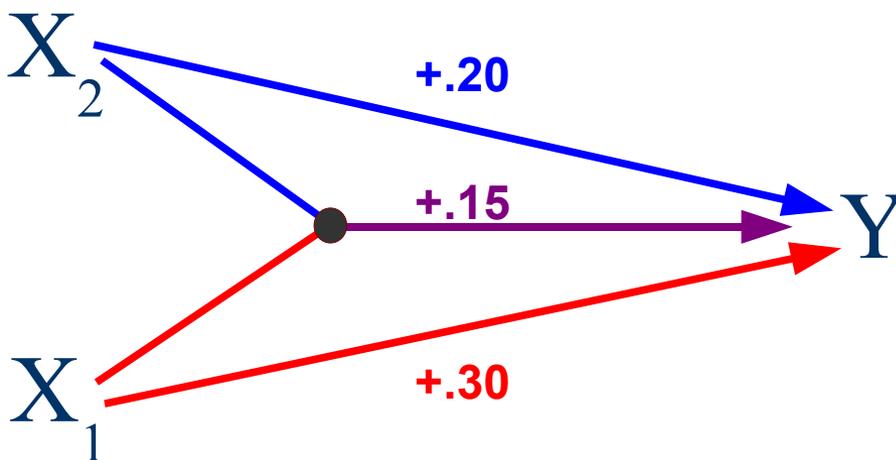
- **Zusammenhangsanalysen III**
  - **Interaktionseffekte im Regressionsmodell**
- **Probeklausur**
- **Besprechung der Lösung**

# Interaktionseffekte im Regressionsmodell

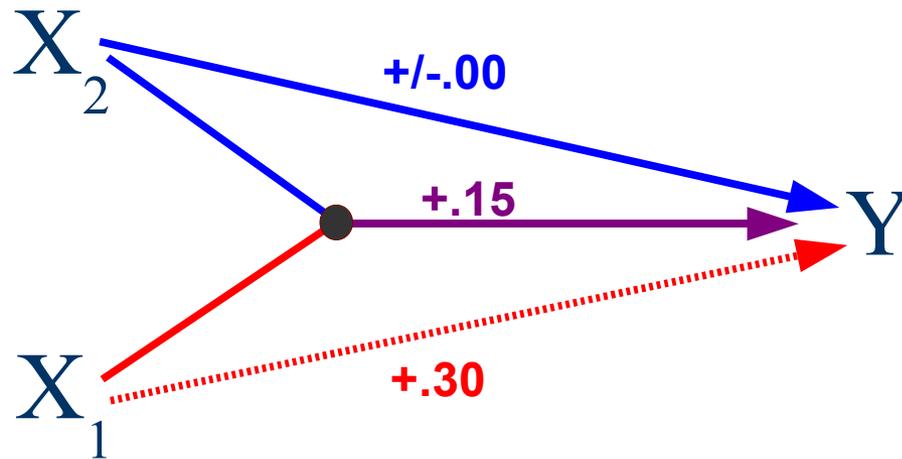
# Moderator oder Interaktion?



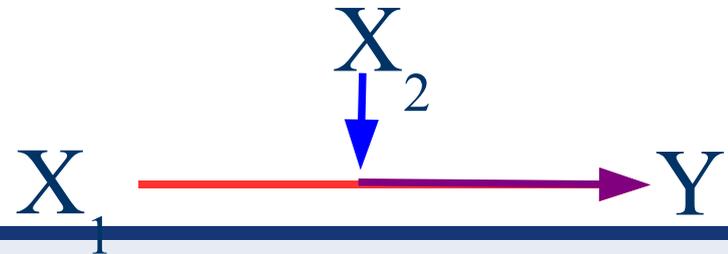
$X_2$  moderiert den Effekt von  $X_1$  auf  $Y$



- $X_1$  und  $X_2$  haben jeweils einen direkten Effekt auf  $Y$ .
- Gleichzeitig interagieren miteinander bei der Erklärung von  $Y$ .



- Ob in diesem Fall eine Moderation oder eine Interaktion vorliegt, ist nur mit experimentellen Daten zu entscheiden. Bei einem reinen Moderationseffekt sollte der Moderator  $X_2$  keinen Effekt auf  $Y$  ausüben, wenn  $X_1 = 0$  ist.



# Wie interagieren Variablen?

- Welche Zusammenhänge lassen sich im Hinblick der Wirkung von Alter und Bildung auf Einkommen erwarten?

1. Je älter eine Person ist, umso höher das Einkommen. **direkte Effekte**
2. Je höher die Bildung, umso höher das Einkommen. **Effekte**

3. Jüngere Akademiker erreichen erst mit einigen Jahren Berufserfahrung das Einkommensniveau von weniger gebildeten Gleichaltrigen.

4. Es existieren eine Reihe von (neueren) Berufen, die nicht über klassische Bildung erreicht werden, aber ein höheres Einkommen ermöglichen. Für junge gebildete Personen im Schnitt nachteilig und für ältere gebildete ein Vorteil im Hinblick auf das Einkommen.

**Hinweise für mögliche Interaktionseffekte**

# Regressionsmodell mit Interaktionseffekt

- `recode v173 (7=.x) (99=.z) (6=.y)`
- `regr v388 v154 v173 , beta`
  
- `egen reg_miss = rowmiss(v388 v154 v173)`
- `sum v173 if reg_miss ==0`
- `gen zv173 = v173 - r(mean) if reg_miss==0`
- `gen zv154 = v154 - r(mean) if reg_miss==0`
- `gen zzalter_bild = zv154*zv173`
- `regr v388 zv154 zv173 zzalter_bild`

# Ergebnisse ohne Interaktion

```
. regress v388 v154 v173 , beta
```

source	SS	df	MS			
Model	332003688	2	166001844	Number of obs =	2759	
Residual	2.6500e+09	2756	961544.752	F( 2, 2756) =	172.64	
				Prob > F	= 0.0000	
				R-squared	= 0.1113	
				Adj R-squared	= 0.1107	
Total	2.9820e+09	2758	1081225.9	Root MSE	= 980.58	

v388	Coef.	Std. Err.	t	P> t	Beta
v154	8.082215	1.105775	7.31	0.000	.1367036
v173	302.1215	16.37842	18.45	0.000	.3450053
_cons	6.94189	88.39754	0.08	0.937	.

# Ergebnisse mit Interaktion

```
. regress v388 zv154 zv173 zzalter_bild
```

Source	SS	df	MS
Model	405898712	3	135299571
Residual	2.5761e+09	2755	935071.619
Total	2.9820e+09	2758	1081225.9

Number of obs = 2759  
 F( 3, 2755) = 144.69  
 Prob > F = 0.0000  
**R-squared = 0.1361**  
**Adj R-squared = 0.1352**  
 Root MSE = 966.99

v388	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
zv154	<b>8.795363</b>	1.093393	8.04	0.000	6.65141	10.93932
zv173	-80.94059	46.01826	-1.76	0.079	-171.1744	9.293174
zzalter_bild	7.999952	.8999165	8.89	0.000	6.235373	9.764531
_cons	974.7247	56.3726	17.29	0.000	864.1879	1085.262

# Interpretation der Effekte

## Regressionsgleichung:

- $\text{Einkommen} = b_0 + b_1 \cdot \text{Alter} + b_2 \cdot \text{Bildung} + b_3 \cdot \text{Alter} \cdot \text{Bildung} + \text{Fehler}$

## Umgeformte Regressionsgleichung:

- $\text{Einkommen} = (b_0 + b_2 \cdot \text{Bildung}) + (b_1 + b_3 \cdot \text{Bildung}) \cdot \text{Alter} + \text{Fehler}$
- $\text{Einkommen} = (974,7 - 81 \cdot 0) + (8,8 + 8 \cdot 0) \cdot 20 = 1150,70$

v388	Coef.
zv154	8.795363
zv173	-80.94059
zalter_bild	7.999952
_cons	974.7247

# Tabellenkalkulation

Auswirkung des <u>Interaktionseffekts</u>			
Einkommen	Bildung	Alter	
992,70	-2	20	
1071,70	-1	20	
1150,70	0	20	
1229,70	1	20	
1308,70	2	20	
704,70	-2	60	
1103,70	-1	60	
1502,70	0	60	
1901,70	1	60	
2300,70	2	60	

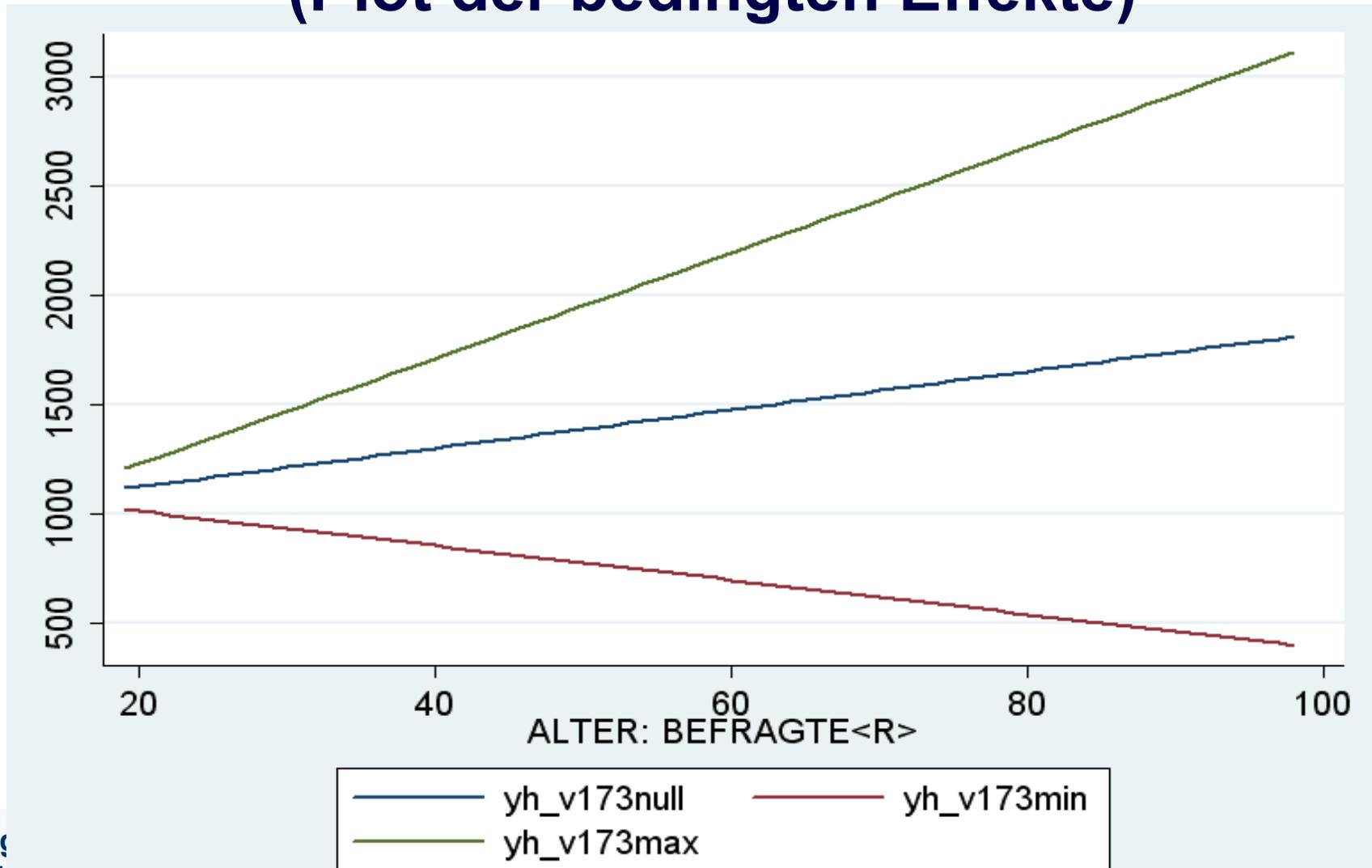
  

zv173	Freq.	Percent	Cum.
-2.081551	67	2.43	2.43
-1.081551	1,028	37.26	39.69
-.0815513	889	32.22	71.91
.9184487	163	5.91	77.82
1.918449	612	22.18	100.00
Total	2,759	100.00	

# Graphische Darstellung

- `gen yh_v173null = _b[_cons] + _b[zv154] *  
zv154`
- `sum zv173`
- `gen yh_v173min = _b[_cons] +  
_b[zv154]*zv154 + _b[zv173]*r(min) +  
_b[zzalter_bild]*zv154*r(min)`
- `gen yh_v173max = _b[_cons] +  
_b[zv154]*zv154 + _b[zv173]*r(max) +  
_b[zzalter_bild]*zv154*r(max)`
- `graph twoway line yh_v173* v154, sort`

# Auswirkung der Interaktion (Plot der bedingten Effekte)



# Probeklausur

# Aufgaben

1. Öffnen Sie den ALLBUS 2008 Datensatz und verwenden Sie den Do-File für die Missings. (5P)
2. Geben Sie an, welche Variablen den Begriff "Bundestag" enthalten. (5P)
3. Versehen Sie die Variable zur Schulbildung des Befragten mit den fehlenden Missings. (10P)
4. Lassen Sie sich die Häufigkeiten für die Bildungsvariable ausgeben. (5P)
5. Berechnen einen additiven Index für die Variablen zu Politikgesprächen im privaten Umfeld. (10P)
6. Weisen Sie der neuen Variable für die Extremwerte Wertelabels zu. (5P)
7. Lassen Sie sich eine bivariate Tabelle für die Variablen Bildung und politisches Interesse des Befragten ausgeben, die auch den Tau\_b-Koeffizienten liefert. (15P)
8. Lassen Sie sich unter Berücksichtigung der korrekten Gewichtung getrennt nach Bundesländern Box-Plots für den Body-Mass-Index (BMI) ausgeben. (20P)
9. Erstellen Sie ein Kreisdiagramm für die Sonntagsfrage. (5P)
10. Geben Sie an, welche Freizeitaktivität in einem multiplen Regressionmodell zur Erklärung des BMI den stärksten und welche den schwächsten signifikanten Effekt besitzt (20P)