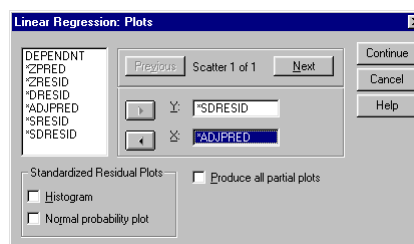
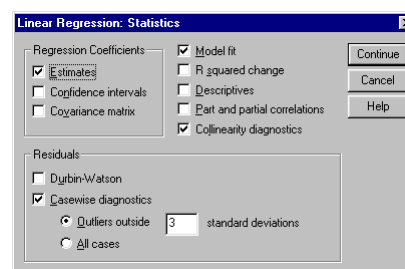
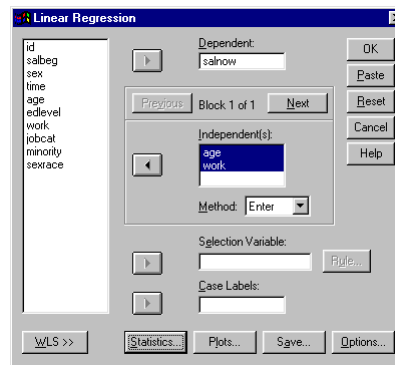


# Aðfallsgreining í SPSS

- Opnaðu gagnaskrá í SPSS, ef þú hefur ekki gert það nú þegar.
  - Ég opna skrá sem heitir Bank.sav með ýmsar upplýsingar um bankastarfsmenn.
- Farðu í valröndina og veldu Statistics /Regression og síðan Linear svo Linear Regression valmyndin birtist.
  - Valmyndin gerir kleift að tilgreina háðu breytuna, óháðu breytur og hvernig þær koma inn í líkanið. Einnig má skilgreina ýmsar samantektartölur (Statistics) og greinandi myndrit. Að síðustu má vista spágildi, leif og ýmsa greiningarstuðla.
- Veldu fylgibreytu og frumbreytur með því að skyggja viðkomandi breytur í listareitnum og færa í viðeigandi textareiti með því að smella á örvatakkanu.
  - Ég vel salnow (núverandi laun) sem fylgibreytu og Age (lífaldur) og Work (starfsaldur) sem frumbreytur.
  - Í reitnum Method má velja hvernig frumbreyturnar koma inn í líkanið. Ég vel að nota Enter sem hentar yfirleitt best en einnig er val um aðrar aðferðir sem geta hentað við sérstakar aðstæður.
  - Það má ýta á Next og þá má skilgreina mengi frumbreyta sem annað hvort fara inn í líkanið eða út úr því. Þannig gæti ég ýtt á Next og fengið aðra sams konar valmynd. Þar gæti ég valið Work sem frumbreytu og valið Remove undir Method. Þá myndi forritið byrja á því að setja bæði Age og Work inn í líkanið og birta niðurstöður. Síðan myndi Work vera tekið út og niðurstöður birtar aftur. Hægt er að skilgreina mörg slíkt skref og þannig sjá hvernig líkanið breytist við það að frumbreytur fara inn í og út úr líkaninu.
- Ýttu nú á Statistics hnappinn til að skilgreina þær samantektartölur sem þú vilt fá. Þá opnast Linear Regression: Statistics valmyndin. Þar merkirðu við þær upplýsingar sem þú vilt að forritið gefi þér. Að því loknu smellir þú á Continue.
  - Ég vel að fá hallatölur (Estimates),  $R^2$  (Model fit), þóltölur (Collinearity diagnostics) og upplýsingar um mögulega frávillinga (Casewise diagnostics).
- Í upphafsvalmyndinni (Linear Regression) smellir þú á Plots og færð Linear Regression: Plots valmyndina. Þar skilgreinir þú þau greinandi myndrit sem þú óskar eftir. Í lokin smellir þú á Continue til að komast í upphafsvalmyndina.
  - Ég vel að fá fylgnirit (*scattergraph*) með *Studentized* leif (SResid) á lóðréttu ásnum og leiðréttu spágildi (AdjPred) á láréttu ásnum. Þetta er hið klassíska leifarrit aðfallsgreiningar.
- Í upphafsvalmyndinni ýtir þú á OK til að láta SPSS hefja úrvinnsluna. Þú getur einnig smellt á Poste til að fá skipanirnar yfir í skipanagluggann (*syntax*). Þar geturðu breytt skipunum og keyrt þær þaðan. Hvora aðferðina sem þú notar, þá hefst SPSS handa og birtir niðurstöðurnar í „Output Navigator.“  
Skipanirnar sem við höfum skilgreint í valmyndunum sjást í textareitnum hér til hliðar.



```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA COLLIN TOL
/NOORIGIN
/DEPENDENT salnow
/METHOD=ENTER age work
/SCATTERPLOT=( *SDRESID , *ADJPRED )
/CASEWISE PLOT(ZRESID) OUTLIERS(3) .
```

# Niðurstöður aðfallsgreiningar í SPSS

## Regression

Variables Entered/Removed<sup>a</sup>

Model	Variables Entered	Variables Removed	Method
1	WORK EXPERIE NCE, AGE OF EMPLOY EE <sup>a</sup>		Enter

a. All requested variables entered.

b. Dependent Variable: CURRENT SALARY

Hér tilgreinir SPSS hvaða frumbreytur eru í líkaninu. Við tilgreindum aðeins eitt líkan en með því að nota röð Method=Enter eða Method=Remove skipana má skilgreina röð líkana sem SPSS reiknar.

Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,150 <sup>a</sup>	,022	,018	6767,61

a. Predictors: (Constant), WORK EXPERIENCE, AGE OF EMPLOYEE

b. Dependent Variable: CURRENT SALARY

Hér kemur yfirlit yfir forspárhæfni líkansins. R<sup>2</sup> er lágt og því gefur líkanið ekki góða forspá.

Staðalvilla spágildis gefur dreifingu raungilda í kringum spágildin, þ.e. dreifingu villunnar, sem er annar mælikvarði á forspá.

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4,9E+08	2	2,5E+08	5,399	,005 <sup>a</sup>
	Residual	2,2E+10	471	4,6E+07		
	Total	2,2E+10	473			

a. Predictors: (Constant), WORK EXPERIENCE, AGE OF EMPLOYEE

b. Dependent Variable: CURRENT SALARY

Þetta er hefðbundin dreifigreiningartafla sem gefur marktækt líkansins í heild sinni.

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error				Beta	Tolerance
1	(Constant)	17536,649	1334,655		13,139	,000		
	AGE OF EMPLOYEE	-110,850	44,436	-,191	-2,495	,013	,353	2,833
	WORK EXPERIENCE	44,219	60,098	,056	,736	,462	,353	2,833

a. Dependent Variable: CURRENT SALARY

Hér birtast stuðlar líkansins, þ.e. hallatölurnar, staðalvillur, marktækt hallastuðla og þoltölur.

Collinearity Diagnostics<sup>a</sup>

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	AGE OF EMPLOYEE	WORK EXPERIENCE
1	1	2,640	1,000	,01	,00	,02
	2	,340	2,786	,05	,00	,36
	3	1,945E-02	11,653	,94	,99	,62

a. Dependent Variable: CURRENT SALARY

Við getum horft fram hjá þessari töflu.

**Casewise Diagnostics<sup>a</sup>**

Case Number	Std. Residual	CURRENT SALARY	Predicted Value	Residual
2	4,105	41400	13618,82	27781,18
55	3,040	34500	13924,44	20575,56
56	6,092	54000	12773,88	41226,12
114	3,667	38800	13981,90	24818,10
123	3,228	36250	14407,18	21842,82
132	4,532	44250	13581,43	30668,57
145	3,487	36800	13202,00	23598,00
149	4,110	41500	13684,14	27815,86
177	3,308	36500	14110,94	22389,06
415	3,828	40000	14092,15	25907,85

a. Dependent Variable: CURRENT SALARY

Hér koma upplýsingar um færslur með áberandi hátt tölugildi leifar. Sumir þessara einstaklinga gætu því verið frávillingar (*outliers*).

Það er eftirtektarvert að leifin er ætíð jákvæð og stundum mjög há eða allt upp í sex staðalfrávik frá spágildinu.

**Residuals Statistics<sup>a</sup>**

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	10867,36	15022,18	13767,83	1022,57	474
Std. Predicted Value	-2,836	1,227	,000	1,000	474
Standard Error of Predicted Value	311,88	1462,14	496,80	207,75	474
Adjusted Predicted Value	10834,54	15040,20	13775,85	1011,07	474
Residual	-7505,41	41226,13	-2,45E-12	6753,29	474
Std. Residual	-1,109	6,092	,000	,998	474
Stud. Residual	-1,111	6,106	-,001	1,000	474
Deleted Residual	-7538,67	41415,16	-8,03	6781,30	474
Stud. Deleted Residual	-1,112	6,356	,002	1,009	474
Mahal. Distance	,007	21,081	1,996	3,103	474
Cook's Distance	,000	,057	,001	,003	474
Centered Leverage Value	,000	,045	,004	,007	474

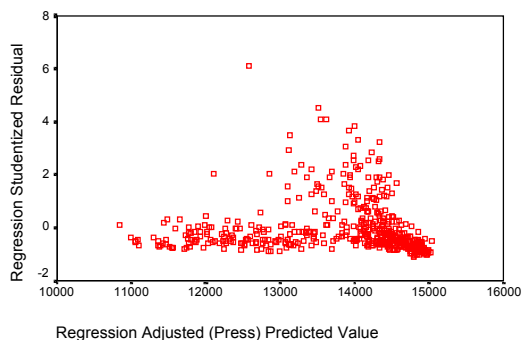
a. Dependent Variable: CURRENT SALARY

Hér kemur yfirlit yfir spágildi, leif og helstu greiningarstuðla. Við fáum lægsta og hæsta gildi, meðaltal og staðalfrávik fyrir alla greiningarstuðla sem SPSS hefur á færi sínu.

Þetta gerir lítið annað en staðfesta það sem við sáum í töflunni hér á undan.

**Scatterplot**

Dependent Variable: CURRENT SALARY



Hér fáum við hefðbundið leifarrit. Ritið er greinilega mjög óeðlilegt sérstaklega þar sem mikil frávik eru í jákvæða átt en lítil neikvæð leif.

Leifin virðist hæst fyrir meðalhá spágildi sem gæti bent til ólínulegra tengsla frumbreyta við fylgibreytu.

Einnig virðist sem leifin sé jákvætt skekkt og því ekki normaldreifð villa.

## Yfirlit yfir Regression skipunina í SPSS

Þetta er stýtt og einfaldað yfirlit. Yfirlit yfir alla möguleika má fá í hjálpinni sem fylgir SPSS.

```
REGRESSION
  [/VARIABLES={varlist      }]
                {(COLLECT)**} Command Syntax
                {ALL        }
  [/DESCRIPTIVES=[DEFAULTS] [MEAN] [STDDEV] [CORR] [COV]
                 [VARIANCE] [XPROD] [SIG] [N] [BADCORR]
                 [ALL] [NONE**]]
  [/SELECT={varname relation value}
  [/MISSING=[{LISTWISE**      }] [INCLUDE]]
            {PAIRWISE        }
            {MEANSUBSTITUTION}
  [/STATISTICS=[DEFAULTS**] [R**] [COEFF**]
               [ANOVA**] [OUTS**]
               [ZPP] [CHA] [CI] [F] [BCOV]
               [SES] [LINE] [HISTORY] [XTX] [COLLIN]
               [END] [TOL] [SELECTION] [ALL]]

  [/{NOORIGIN**}]
    {ORIGIN      }
  /DEPENDENT=varlist
  [/METHOD={STEPWISE [varlist]      }
           {FORWARD [varlist]       }
           {BACKWARD [varlist]      }
           {ENTER [varlist]         }
           {REMOVE varlist          }
           {TEST(varlist) (varlist)...}]
  [/RESIDUALS=[DEFAULTS] [ID(varname)]
              [DURBIN] [{SEPARATE}]
                      {POOLED }
              [HISTOGRAM({ZRESID    })]
                      {tempvarlist}
              [OUTLIERS({ZRESID    })]
                      {tempvarlist}
              [NORMPROB({ZRESID    })]
                      {tempvarlist}
  [/CASEWISE=[DEFAULTS] [{OUTLIERS({ 3 })}]
              {value} }
              {ALL      }
  [/SCATTERPLOT=(varname, varname)]
  [/PARTIALPLOT={ALL      }
                {varlist}]

  [/SAVE=tempvar[(name)] [tempvar[(name)]...] [FITS]]
```

Temporary variables for residuals analysis are:

PRED, ADJPRED, SRESID, MAHAL, RESID, ZPRED, SDRESID, COOK,  
DRESID, ZRESID, SEPRED, LEVER, DFBETA, SDBETA, DFFIT, SDFFIT,  
COVRATIO, MCIN, ICIN.

SAVE FITS saves DFFIT, SDFIT, DFBETA, SDBETA, COVRATIO.

\*\*Default if the subcommand is omitted.