Regression Analysis: Third lecture

Pierre-Jérôme Bergeron
University of Ottawa

September 11, 2014

Overview

Previously in MAT 3375

Read the above in Morgan Freeman's voice*

We saw

- The course in one slide
- The basics (mostly point estimation) of simple linear regression
  - $E(b_1) = \beta_1$
  - $Var(b_1) = \frac{\sigma^2}{S_{XX}}$
  - $E(b_0) = \beta_0$
  - $Var(b_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right)$
  - b_1 and b_0 are maximum likelihood estimators (i.e. they are $\hat{\beta}_1$ and $\hat{\beta}_0$)

* You really should know this American actor, who has done many roles. Here's a short audio bit of him recording Duke Nukem lines.

Today

For the aspiring statistician

- Inference related to simple linear regression
  - For those following the book, it's Chapter 2, already.

MAT 3375 - Regression Analysis

Statistical inference for β_1

- Although we know the expectation and variance of our estimators, b_0^* and b_1, we also need distributions for the estimators to make inference
- Usually we make inference through hypothesis tests and by building confidence intervals
- In order to do this exactly, we'll need to assume that the data are normally distributed, usually accomplished by assuming that $\epsilon_i \sim N(0, \sigma^2)$
- Approximate normality through asymptotics will be discussed later
- Hypothesis testing typically centers around the slope, β_1
- Most common null hypothesis is *simple*: $H_0 : \beta_1 = \beta_{1,0}$

Hypothesis Testing

Is the pattern real or just noise?

- The typical simple linear regression hypothesis of interest is that the explanatory variable (or covariate) is linearly associated with the response variable OR that one can use the covariate to predict the response better than just by using the mean of the response
- For either, we can say that when there is a linear association between the two variables that $\beta_1 \neq 0$
- $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$
- Need
    - Test statistic
    - Distribution of test statistic under the null hypothesis

$t-$**test for** β_1 **and** β_0

t, Earl Grey, hot...

- Can test using our estimator $\hat{\beta}_1 = b_1$ and properties of it
- Remember

$$\hat{\beta}_1 = b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- The ϵ_i are independent normals, so y_i are independent normals
- The x_i's are constant
- Therefore, $\hat{\beta}_1$ is a weighted sum of normal random variables (and so is also normal)

$t-$test for β_1 and β_0

Read *The Lady Tasting Tea*

- Know the expected value (β_1) and variance (σ^2/S_{XX}) of $\hat{\beta}_1$, so

$$\hat{\beta}_1 \sim \text{Normal}(\beta_1, \sigma^2/S_{XX})$$

- Let $SS_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (b_0^* + (x_i - \bar{x})b_1))^2$
- Can show that $SS_{Res}/\sigma^2 \sim \chi^2_{n-2}$ (will prove when we move to multiple linear regression)
- Let

$$s^2 = \frac{SS_{Res}}{n-2}$$

$t-$test for β_1 and β_0
It's all very mechanical

- Basic statistical inference and results from before about SS_{Res} yields under $H_0 : \beta_1 = 0$:

$$t = \frac{b_1 - 0}{\sigma/\sqrt{S_{XX}}} \times \frac{\sigma}{s} = \frac{b_1 - 0}{s/\sqrt{S_{XX}}} \sim t_{n-2}$$

- Can adjust to test $H_0 : \beta_1 = \beta_{1,0}$

$$t = \frac{b_1 - \beta_{1,0}}{s/\sqrt{S_{XX}}} \sim t_{n-2}$$

  - $H_0 : \beta_1 = 0$ translates to x and y are un(cor)related, i.e. y does not depend on x.
  - $H_0 : \beta_1 = \beta_{1,0}$ assumes *prior knowledge* of the linear relationship between x and y at the value of $\beta_{1,0}$

Inference for the intercept

If anyone cares

- Like $\hat{\beta}_1$, $b_0 = \hat{\beta}_0$ is a linear combination of independent normals,

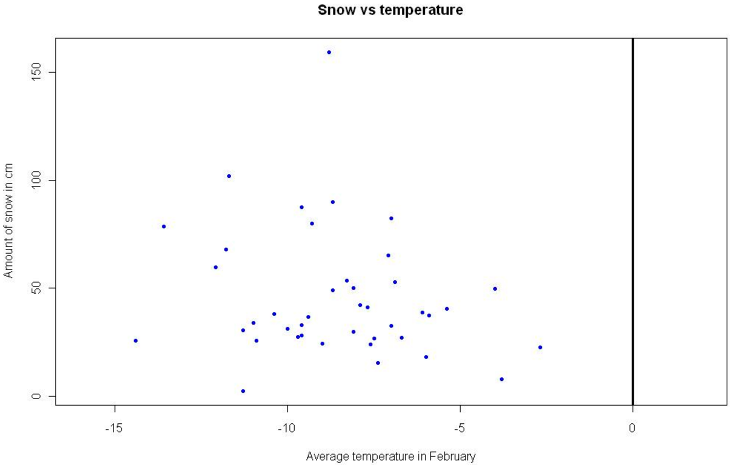
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- We also know the mean and the variance of $\hat{\beta}_0$, so for $H_0 : \beta_0 = \beta_{0,0}$ we can use:

$$t = \frac{b_0 - \beta_{0,0}}{s\sqrt{(1/n) + (\bar{x}^2/S_{XX})}} \sim t_{n-2}$$

Snow data

Raw data in colours!



Snow data

Using SAS

```
data mat3375.snow;
input year temp snow;
datalines;
 1971  -8.8  159.5
 ...
 2011  -8.1   50.0
;
run;
ods graphics on;
proc REG data=mat3375.snow;
  model snow = temp / clb;
run;
ods graphics off;
```

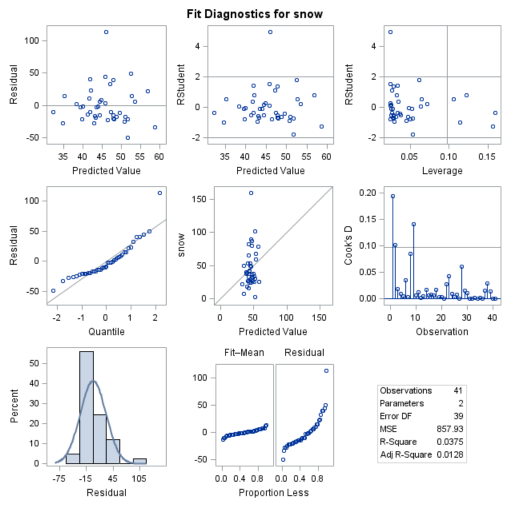
Some of the output
Your grandma's birthday omitted

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	26.20675	16.37079	1.60	0.1175	-6.90630	59.31980
temp	1	-2.25511	1.83035	-1.23	0.2253	-5.95734	1.44713

- SAS gives way too much output
- It gives too many decimals by default too
- ODS Graphics now standard with SAS 9.3
- Without it, you'll get less fancy graphs

Snow data with SAS

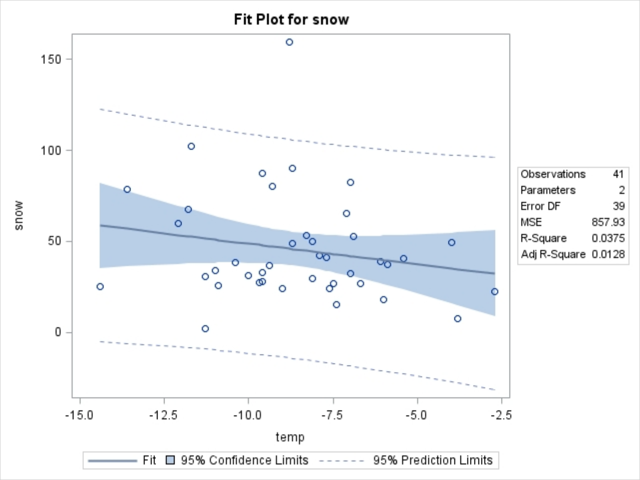
ODS Graphics Overload



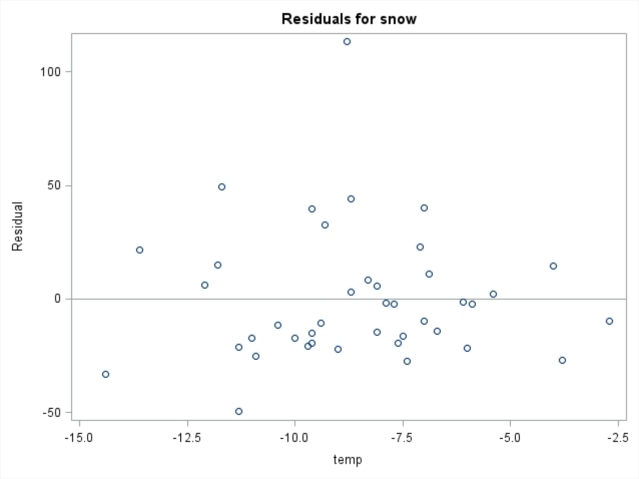
Observations	41
Parameters	2
Error DF	39
MSE	857.93
R-Square	0.0375
Adj R-Square	0.0128

Snow data with SAS

We'll cover the rest later



More SAS output
Residual plot



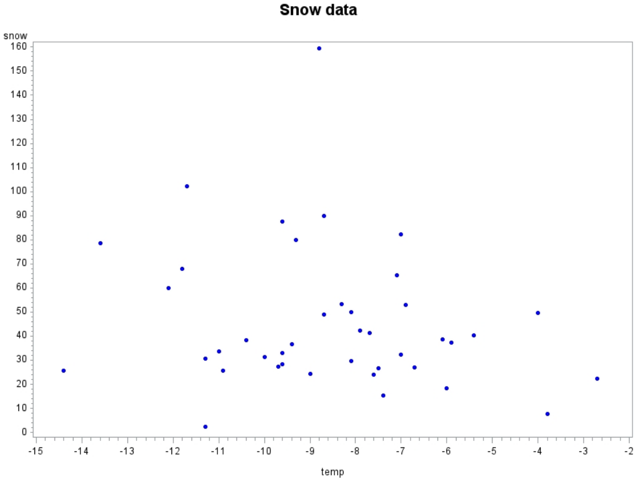
Getting less output

Your basic scatterplot

```
 symbol1 v=dot h=.8 c=blue;
TITLE 'Snow data';
proc gplot data = MAT3375.snow;
plot snow*temp;
run;
```

Less SAS output

Scatter plot



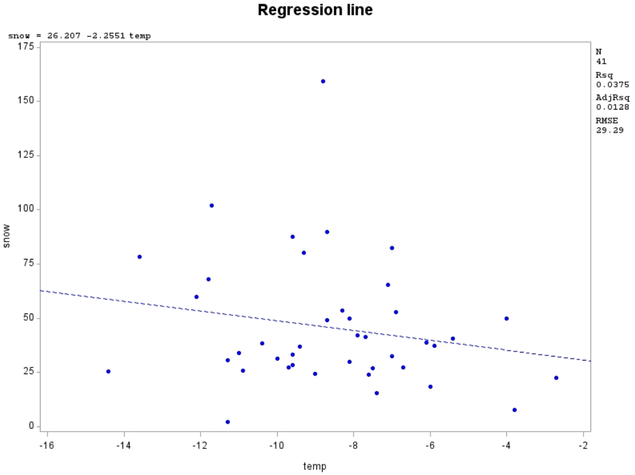
Just the regression line

Well, and a little bit more

```
 TITLE 'Regression line';
proc reg data = MAT3375.snow;
model snow = temp;
plot snow*temp;
run;
```

Less SAS output

Regression plot



Partitioning variability

Model (regression) vs residuals

- Consider the regression problem as a paritioning of the variability of the response, i.e. $\sum_{i=1}^{n}(y_i - \bar{y})^2$
- Simple algebra shows that:

Partitioning variability

Model (regression) vs residuals

- Consider the regression problem as a paritioning of the variability of the response, i.e. $\sum_{i=1}^{n}(y_i - \bar{y})^2$
- Simple algebra shows that:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$$

Partitioning variability

Model (regression) vs residuals

- Consider the regression problem as a paritioning of the variability of the response, i.e. $\sum_{i=1}^{n}(y_i - \bar{y})^2$
- Simple algebra shows that:

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)
\end{aligned}$$

Partitioning variability

Model (regression) vs residuals

- Consider the regression problem as a paritioning of the variability of the response, i.e. $\sum_{i=1}^{n}(y_i - \bar{y})^2$
- Simple algebra shows that:

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2
\end{aligned}
$$

Partitioning variability

Exercise: do this on paper at home

- Showing $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$....

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = [\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i)] - \bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

Partitioning variability

Exercise: do this on paper at home

- Showing $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$....

$$\begin{aligned}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= [\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i)] - \bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i) \\ &= \sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i)\end{aligned}$$

Partitioning variability

Exercise: do this on paper at home

- Showing $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$....

$$
\begin{aligned}
\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= [\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i)] - \bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}(b_0 + b_1 x_i)(y_i - \hat{y}_i)
\end{aligned}
$$

Partitioning variability

Exercise: do this on paper at home

- Showing $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$....

$$
\begin{aligned}
\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= [\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i)] - \bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}(b_0 + b_1 x_i)(y_i - \hat{y}_i) \\
&= b_1\sum_{i=1}^{n}x_i(y_i - \hat{y}_i)
\end{aligned}
$$

Partitioning variability

Exercise: do this on paper at home

- Showing $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$....

$$
\begin{aligned}
\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= [\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i)] - \bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}(b_0 + b_1 x_i)(y_i - \hat{y}_i) \\
&= b_1\sum_{i=1}^{n}x_i(y_i - \hat{y}_i) \\
&= b_1 \cdot 0 \text{ by Homework 1}
\end{aligned}
$$

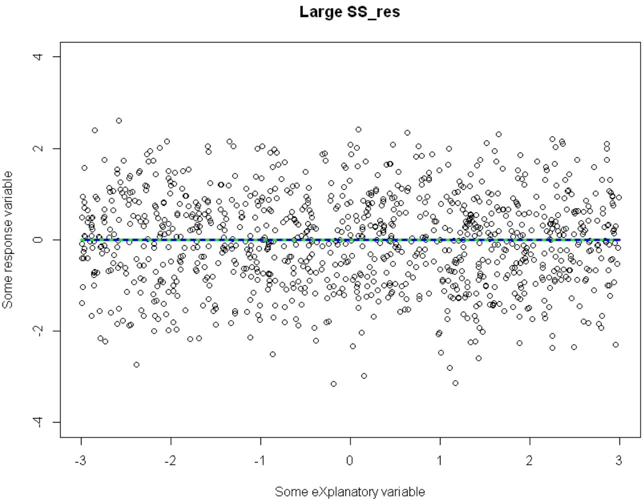
Partitioning variability

Not a musical term

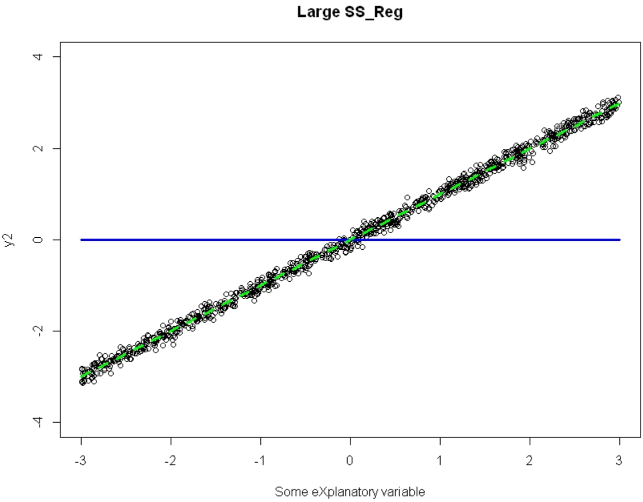
- We now have shown that the overall variability in $\sum_{i=1}^{n}(y_i - \bar{y})^2$ can be partitioned into $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ [SS_{Reg}] and $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ [SS_{Res}]
- It can be described as
  - the variability of y (the response) explained by the model (i.e. by the variability of X) that's the SS_{Reg} part
  - the variability of y left in the residuals (SS_{Res}), i.e. the variability not explained by the model
- We can also see this graphically

Larger SS_{Res}

Lots of variability in the residuals



Larger SS_{Reg}
y varies with x, i.e. strong correlation



Partitioning the variability

A preview

- Intuitively, our SS_{Reg} and SS_{Res} will be very important for determining the *significance* and *extent* of our model fit
- Before we can actually test any hypotheses, we also will need to talk about partioning the degrees of freedom for our dataset

Degrees of freedom
They should add up

- (at least) Two ways to think about degrees of freedom
  - Number of independent observations minus the number of parameters estimated
  - Sample size reduced by the number of restrictions on the model residuals

Partitioning degrees of freedom

Insert an Yvon Deschamps joke here

- $SS_{Tot}(\sum_{i=1}^{n}(y_i - \bar{y})^2)$ has $(n-1)$ independent pieces of information because $\sum_{i=1}^{n} y_i/n = \bar{y}$
- The $\hat{y}_i$ are constrained to lie on a line, i.e. there are only 2 independent pieces of information (the slope and the intercept)
- Consider the centered model, then $b_0 = \bar{y}$ (so we no longer have 2 independent ways to make the y_i deviate from the mean)
- $SS_{Reg}(\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2)$ should have 1 degree of freedom, which means there is only one piece of independent information in the sum (the slope, because $b_0^* = \bar{y}$)
- $SS_{Tot} = SS_{Reg} + SS_{Res}$
- $df_{Tot} = df_{Reg} + df_{Res}$

Hypothesis Testing
Old-fashioned stats

- What do we have available for testing?
- Could use estimator of β_1, could also use partitioning of total variance
- If using partitioning, could consider
  1. SS_{Reg} and SS_{Tot}
  2. SS_{Reg} and SS_{Res}

Facts about SS_{Reg} and SS_{Res}

Please do your own fact checking at home

- SS_{Reg} and SS_{Res} are independent
- $SS_{Reg}/\sigma^2 \sim \chi^2_1$ under the null hypothesis
- $SS_{Res}/\sigma^2 \sim \chi^2_{n-2}$ under the null hypothesis
- $F = \frac{SS_{Reg}/1}{SS_{Res}/(n-2)}$ therefore has an $\mathcal{F}$ distribution with 1 and $n-2$ degrees of freedom ($\mathcal{F}_{1,n-2}$)
- Proofs postponed until later

Two interpretations

There's always more than one way to see things

- Comparing the relative size of deviation of the line from the mean and the observations from the line
- Comparing how much SS_{Reg} deviates from σ^2 (why?)

Second interpretation

Depth perception

- $E(SS_{Res}/(n-2)) = E(s^2) = \sigma^2$

Second interpretation

Depth perception

- $E(SS_{Res}/(n-2)) = E(s^2) = \sigma^2$

$$E(SS_{Reg}) = E(\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2)$$

Second interpretation

Depth perception

- $E(SS_{Res}/(n-2)) = E(s^2) = \sigma^2$

$$
\begin{aligned}
E(SS_{Reg}) &= E(\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^{n} E((\hat{y}_i - \bar{y})^2)
\end{aligned}
$$

Second interpretation

Depth perception

- $E(SS_{Res}/(n-2)) = E(s^2) = \sigma^2$

$$
\begin{aligned}
E(SS_{Reg}) &= E(\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^{n} E((\hat{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^{n} E((b_0^* + b_1(x_i - \bar{x}) - \bar{y})^2)
\end{aligned}
$$

Second interpretation
Depth perception

- $E(SS_{Res}/(n-2)) = E(s^2) = \sigma^2$

$$
\begin{aligned}
E(SS_{Reg}) &= E(\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^{n} E((\hat{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^{n} E((b_0^* + b_1(x_i - \bar{x}) - \bar{y})^2) \\
&= \sum_{i=1}^{n} E((b_1(x_i - \bar{x}))^2)
\end{aligned}
$$

Second interpretation

Going deeper

$$\sum_{i=1}^{n} E((b_1(x_i - \bar{x}))^2) = S_{XX} \times E(b_1^2)$$

Second interpretation

Going deeper

$$\sum_{i=1}^{n} E((b_1(x_i - \bar{x}))^2) = S_{XX} \times E(b_1^2)$$

$$= S_{XX} \times [Var(b_1) + E(b_1)^2]$$

Second interpretation

Going deeper

$$
\begin{aligned}
\sum_{i=1}^{n} E((b_1(x_i - \bar{x}))^2) &= S_{XX} \times E(b_1^2) \\
&= S_{XX} \times [Var(b_1) + E(b_1)^2] \\
&= S_{XX} \times (\sigma^2 / S_{XX} + \beta_1^2)
\end{aligned}
$$

Second interpretation

Going deeper

$$\begin{aligned}
\sum_{i=1}^{n} E((b_1(x_i - \bar{x}))^2) &= S_{XX} \times E(b_1^2) \\
&= S_{XX} \times [Var(b_1) + E(b_1)^2] \\
&= S_{XX} \times (\sigma^2 / S_{XX} + \beta_1^2) \\
&= \sigma^2 + \beta_1^2 S_{XX}
\end{aligned}$$

Second interpretation

Going deeper

$$\begin{aligned}
\sum_{i=1}^{n} E((b_1(x_i - \bar{x}))^2) &= S_{XX} \times E(b_1^2) \\
&= S_{XX} \times [Var(b_1) + E(b_1)^2] \\
&= S_{XX} \times (\sigma^2/S_{XX} + \beta_1^2) \\
&= \sigma^2 + \beta_1^2 S_{XX}
\end{aligned}$$

- If $\beta_1 = 0$ it's another estimator of σ^2, otherwise it's (stochastically) *bigger*

Organizing information
A preview of MAT 3378

- Often ANOVA (ANalysis Of VAriance) results are presented in the form of a table

Source	SS	df	Mean Square (MS)	F
Reg	SS_{Reg}	1	$SS_{Reg}/1$	$F = MS_{Reg}/MS_{Res}$
Res	SS_{Res}	$n-2$	$s^2 = SS_{Res}/(n-2)$	
Tot	SS_{Tot}	$n-1$	$SS_{Tot}/(n-1)$	

Likelihood ratio test

LRT for short

- $F-$test is a test of the alternative model with $\beta_1 \neq 0$ vs. the null model with $\beta_1 = 0$
- Can also view the $F-$test as a version of a likelihood ratio test
- More on this later, maybe (perhaps in MAT 4175, which will be taught by Farid Elktaibi, this winter term)

Snow data SAS output

That was left out earlier

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1302.31656	1302.31656	1.52	0.2253
Error	39	33459	857.92933		
Corrected Total	40	34762			

Root MSE	29.29043	R-Square	0.0375
Dependent Mean	45.57317	Adj R-Sq	0.0128
Coeff Var	64.27122		

Snow data SAS output

BTW: Standard error (i.e. $\hat{\sigma}$)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1302.31656	1302.31656	1.52	0.2253
Error	39	33459	857.92933		
Corrected Total	40	34762			

$\hat{\sigma}$

Root MSE	29.29043	R-Square	0.0375
Dependent Mean	45.57317	Adj R-Sq	0.0128
Coeff Var	64.27122		

Need a break?

Had enough?

- Me too, time to call it a day
- If there is still time left for the class (doubtful?) your prof is there to answer questions

Next lecture

A preview

- More regression means more SAS
  - By which I mean: it's a crash course in basic SAS
  - Everything else you need will be covered as we go along with the course
- Also, assignment # 1 is due in 2 weeks