## ACTEX Learning Flashcards

Learning & Memorizing Key Topics and Formulas

# Solution Spring 2019 Edition

Runhuan Feng, Ph.D., FSA, CERA, Daniël Linders, Ph.D. Ambrose Lo, Ph.D., FSA, CERA Copyright © 2019, ACTEX Learning, a division of SRBooks Inc.

No portion may be reproduced or transmitted in any part or by any means without the permission of the publisher.

ISBN 978-1-63588-708-2

Printed in the United States of America.

ACTEX is committed to making continuous improvements to our study material. We thus invite you to provide us with a critique of these flashcards.

Publication: ACTEX SOA SRM Flashcards, Spring 2019 Edition

In preparing for my exam I found this material: (Check one)

\_\_\_\_Very Good \_\_\_\_Good \_\_\_\_Satisfactory \_\_\_Unsatisfactory

I found the following helpful:

I found the following problems: (Please be specific as to section, specific item, and/or page number)

Please continue on the other side of this card.

#### To improve these flashcards I would:

Name:	
Address: _	
Phone:	
E-mail:	

(Please provide this information in case clarification is needed.)

Send to: Stephen Camilli ACTEX Learning P.O. Box 715 New Hartford, CT 06057

Or visit our website at www.ActexMadRiver.com to complete the survey on-line. Click on the "Send Us Feedback" link to access the online version. You can also e-mail your comments to Support@ActexMadRiver.com.

### Preface

This set of flashcards is meant to complement the *ACTEX Study Manual for SOA Exam SRM (Statistics for Risk Modeling)*. Fully revised in response to the May 2019 edition of the SRM study manual, these flashcards provide a concise summary of the SRM exam material in a readable and presentationoriented format with a view to maximizing retention. Important formulas are displayed to facilitate identification and memorization. Suggestions are given as to which formulas, in our opinion, must be memorized, which formulas are important but can be easily deduced from other results, and which formulas are of secondary importance. The flashcards are particularly suitable for last-minute review—don't forget to take them with you on your way to the CBT exam center!

It should be noted, however, that these flashcards add value to, but are no substitute for reading the SRM study manual. Examples and problems, which are key to exam success, are not included or discussed in these flashcards. We suggest that you first read the manual carefully, go over the in-text examples and (most of the) end-of-chapter problems, then use the flashcards as a means to review what you have learned and to ensure that you have mastered all of the key concepts.

As with the SRM study manual, we would be extremely grateful if you could share your comments and suggestions on these flashcards with us and bring to our attention any potential errors. Please direct your comments and questions to ambrose-lo@uiowa.edu. The authors will try their best to respond to any inquiries as soon as possible and an ongoing list of updates will be maintained online at https://sites.google.com/site/ambroseloyp/publications/SRM.

We wish you the best of luck with your SRM exam!

Runhuan Feng Daniël Linders Ambrose Lo February 2019

## Part I Regression Models

## Chapter 1 Simple Linear Regression

#### 1.1 Basics

• Simple linear regression (SLR) model equation: An approximately linear relationship between y and x:



where

y is the response variable (a.k.a. dependent variable), x is the explanatory variable (a.k.a. predictors, features),  $\beta_0$  (intercept) and  $\beta_1$  (slope) are regression coefficients,  $\varepsilon$  is the random error term.

In the above model, we say that y is regressed on x (denoted  $y \sim x$ ).

• **Defining property of SLR:** There is only one explanatory variable, namely, *x*.

#### • Model assumptions:

- A1. The  $y_i$ 's are realizations of random variables, while the  $x_i$ 's are nonrandom.
- A2.  $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$  are independent with

$$\mathbb{E}[\varepsilon_i] = 0$$
 and  $\operatorname{Var}(\varepsilon_i) = \sigma^2$ 

for all i = 1, 2, ..., n.

Almost always further assume that  $\varepsilon_i$ 's are normally distributed, i.e.,

 $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ 

#### 1.2 Model Fitting by Least Squares Method

• Idea of least squares method: Choose  $\beta_0$  and  $\beta_1$  to make the sum of squares



the "least."

• Least squares estimates (LSEs):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where

$$\triangleright \ S_{xy} := \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y}$$
  
$$\triangleright \ S_{xx} := \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n \bar{x}^2$$

(Suggestion: Remember the formulas for  $\hat{\beta}_0$  and  $\hat{\beta}_1$ !)

ACTEX Learning © 2019

#### • How can the calculation of LSEs be tested?

- Case 1. Given the raw data  $\{(x_i, y_i)\}_{i=1}^n$  with a relatively small n(e.g.,  $n \leq 10$ ) Enter the data into your financial calculator and read the output from its statistics functions.
- Case 2. Given summarized data in the form of various sums, e.g.,

$$\sum_{i=1}^{n} x_{i}, \qquad \sum_{i=1}^{n} y_{i}, \qquad \sum_{i=1}^{n} x_{i}^{2}, \qquad \sum_{i=1}^{n} y_{i}^{2}, \qquad \sum_{i=1}^{n} x_{i}y_{i}.$$

Expand the products in the two sums that appear in  $\hat{\beta}_1$  and use the alternative form

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}.$$

• An alternative formula for  $\hat{\beta}_1$  in terms of sample correlation:

$$\hat{\beta}_1 = r \times \frac{s_y}{s_x}, \qquad \left( \text{Warning: Not } r \times \frac{s_x}{s_y}! \right)$$

where

 $\triangleright$   $s_x$  and  $s_y$  are the sample standard deviations of x and y  $\triangleright$  r is the sample correlation coefficient between x and y

• Application of this formula: Slope estimates when regressing y on x and regressing x on y are related via

$$\hat{\beta}_1^{y \sim x} \times \hat{\beta}_1^{x \sim y} = \underbrace{r^2 = R^2}_{\text{see Sect. 1.3}}$$

• Fitted values and residuals: Given  $\hat{\beta}_0$  and  $\hat{\beta}_1$ , we can compute:

1. The fitted value (a.k.a. predicted value)  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 

 $\,\vartriangleright\,$  Mnemonic: Obtained from the model equation by

$$\beta_0 \to \hat{\beta}_0, \qquad \beta_1 \to \hat{\beta}_1, \qquad \varepsilon_i \to 0$$

 $\triangleright$  Ideally,  $\hat{y}_i$  should be close to  $y_i$ .

- 2. The residual  $e_i = y_i \hat{y}_i$ 
  - $\triangleright$  Memory alert: Not  $\hat{y}_i y_i!$
  - $\triangleright$  Completely different from  $\varepsilon_i$ , which is unobservable and which  $e_i$  serves to approximate.

• Graphical illustration of fitted regression line and definitions of fitted value and residual:



ACTEX Learning  $\bigcirc 2019$ 

#### • Sum-to-zero constraints on residuals:

- 1.  $\sum_{i=1}^{n} e_i = 0$ , provided that  $\beta_0$  is included in the model Meaning: The residuals offset one another to produce a zero sum; they are negatively correlated.
- 2.  $\sum_{i=1}^{n} x_i e_i = 0$ Meaning: The residuals and the explanatory variable values are uncorrelated.

Mnemonic:  $\hat{\beta}_0$  and  $\hat{\beta}_1$  satisfy

$$\frac{\partial}{\partial \beta_0} SS(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^n [\underbrace{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}_{e_i}] = 0,$$
$$\frac{\partial}{\partial \beta_1} SS(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^n x_i [\underbrace{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}_{e_i}] = 0.$$

ACTEX Learning © 2019

#### 1.3 Assessing Goodness of Fit of the Model

•	Three	kinds	of	$\mathbf{sums}$	of	squares:
---	-------	-------	----	-----------------	----	----------

Sum of Squares	Abbrev.	Def.	What Does It Measure?
$\underline{\mathrm{T}}\mathrm{otal}\;\mathrm{SS}$	TSS	Variation of response values about $\bar{y}$	Amount of variability inher- ent in the response prior to performing regression
<u>R</u> esidual SS or Error SS	RSS	Variation of response values about fitted regression line	<ul> <li>Goodness of fit of the SLR model (the lower, the better)</li> <li>Amount of variability of response left unexplained even after introduction of x</li> </ul>
<u>Reg</u> ression SS	Reg SS	Variation explained by SLR (or the knowledge of $x$ )	How effective SLR model is in explaining the variation in $\boldsymbol{y}$

ACTEX Learning  $\bigodot 2019$ 

CHAPTER 1. SIMPLE LINEAR REGRESSION

• ANOVA identity:

$$\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{RSS}} + \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{\text{Reg SS}}.$$

• Coefficient of determination:

$$\triangleright$$
 Definition:  $R^2 = \frac{\text{Reg SS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$ 

- ▷ Measures the proportion of variation of response (about its mean) explained by the SLR model
- $\triangleright$  The higher, the better
- Specialized formulas for Reg SS and  $R^2$  under SLR:

$$\triangleright \operatorname{Reg} SS = \hat{\beta}_1^2 S_{xx}$$
  
$$\triangleright R^2 = r^2 = \operatorname{Corr}(x, y)^2 \text{ (square of correlation between } x \text{ and } y)$$

#### • ANOVA table:

Source	Sum of Squares	df	Mean Square	F-value
Regression	Reg SS	1	${\rm Reg}~{\rm SS}/1$	?
Error	RSS	n-2	$s^2 = \mathrm{RSS}/(n-2)$	
Total	TSS	n-1		

Structure:

- $\triangleright$  Different sources of variation in y
- $\triangleright$  Some "informal" rules for counting df:
  - Reg SS has 1 df because of one explanatory variable
  - RSS has 2 df subtracted from n because of two parameters,  $\beta_0$  and  $\beta_1$
- $\triangleright$  Dividing an SS by its *df* results in a *mean square* (MS).

• Mean square error:

$$s^{2} = \frac{\text{RSS}}{\text{df of RSS}} = \frac{\sum_{i=1}^{n} e_{i}^{2}}{n-2};$$

 $s=\sqrt{s^2}$  is the residual standard deviation or residual standard error.

- *F*-test:
  - $\triangleright$  Hypotheses:

$$\underbrace{\mathbf{H}_{0}:\,\beta_{1}=0}_{\text{i.i.d. model}} \quad \text{vs.} \quad \underbrace{\mathbf{H}_{a}:\,\beta_{1}\neq 0}_{\text{SLR model}};$$

a test of the significance/usefulness of x in explaining y > F-statistic:

$$F = \frac{\text{Reg SS}/(\text{df of Reg SS})}{\text{RSS}/(\text{df of RSS})} = \frac{\text{Reg SS}/1}{\text{RSS}/(n-2)}$$

- $\triangleright$  Behavior of *F*-statistic:
  - H<sub>0</sub>: Expected value close to one
  - H<sub>a</sub>: Tends to be large

#### • *F*-test: (Cont.)

 $\triangleright$  Going between *F*-statistic and  $R^2$ :

$$F = (n-2) \left(\frac{\text{Reg SS/TSS}}{\text{RSS/TSS}}\right) = (n-2) \left(\frac{R^2}{1-R^2}\right)$$

(Mnemonic: Divide both the numerator and denominator of the F-statistic by TSS to get  $R^2$ .)

#### **1.4** Statistical Inference about $\beta_0$ and $\beta_1$

- Sampling distributions of  $\hat{\beta}_0$  and  $\hat{\beta}_1$ :
  - $\triangleright$  Linear combination formulas:

$$\hat{\beta}_1 = \sum_{i=1}^n w_i y_i, \quad \text{where } w_i = \frac{x_i - \bar{x}}{S_{xx}}$$
$$\hat{\beta}_0 = \sum_{i=1}^n w_{i,0} y_i, \quad \text{where } w_{i,0} = \frac{1}{n} - \bar{x} w_i$$

(Suggestion: Remembering these weights is recommended, but not absolutely essential)

• Sampling distributions of  $\hat{\beta}_0$  and  $\hat{\beta}_1$ : (Cont.)

$$\triangleright$$
 Unbiased:  $\mathbb{E}[\hat{\beta}_j] = \beta_j$  for  $j = 0, 1$ 

 $\triangleright$  Variances:

$$\operatorname{Var}(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \quad \text{and} \quad \operatorname{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

(Suggestion: Remember these two formulas!)

 $\triangleright$  Estimated variances: With  $\sigma^2 \rightarrow s^2$  (MSE),

$$\widehat{\operatorname{Var}}(\hat{\beta}_0) = s^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \quad \text{and} \quad \widehat{\operatorname{Var}}(\hat{\beta}_1) = \frac{s^2}{S_{xx}}$$

- t-test:
  - $\beta_j \neq d$   $\triangleright \ Hypotheses: \ \mathbf{H}_0: \ \beta_j = d \ \text{vs.} \ \mathbf{H}_a: \ \beta_j > d$  $\beta_j < d$
  - $\triangleright$  Important special case: d = 0 (i.e., to test if x is useful)  $\triangleright$  t-statistic:

$$t(\hat{\beta}_j) = \frac{\text{LSE} - \text{hypothesized value}}{\text{standard error of LSE}} = \frac{\hat{\beta}_j - d}{\text{SE}(\hat{\beta}_j)}$$

- $\triangleright$  Null distribution:  $t(\hat{\beta}_j) \stackrel{\mathrm{H}_0}{\sim} t_{n-2}$
- $\triangleright$  Decision rules and p-value:

$H_a$	Decision Rule	<i>p</i> -value ( <i>t</i> is the observed value of $t(\hat{\beta}_j)$ )
$\beta_j \neq d$	$ t(\hat{\beta}_j)  > t_{n-2,\alpha/2}$	$\mathbb{P}( t_{n-2}  >  t ) = 2\mathbb{P}(t_{n-2} >  t )$
$\beta_j > d$	$t(\hat{\beta}_j) > t_{n-2,\alpha}$	$\mathbb{P}(t_{n-2} > t)$
$\beta_j < d$	$t(\hat{\beta}_j) < -t_{n-2,\alpha}$	$\mathbb{P}(t_{n-2} < t)$

ACTEX Learning © 2019

• Confidence intervals (CIs) for  $\beta_0$  and  $\beta_1$ : General structure is

LSE  $\pm t$ -quantile  $\times$  Standard error  $= \hat{\beta}_j \pm t_{n-2,\alpha/2} \times \text{SE}(\hat{\beta}_j).$ 

 $\,\vartriangleright\,$  E.g.:  $\hat{\beta}_1 \pm t_{n-2,\alpha/2} \times \operatorname{SE}(\hat{\beta}_1)$  is the CI for  $\beta_1$ 

- $\triangleright$  Construction requires formulas of  $SE(\hat{\beta}_0)$  and  $SE(\hat{\beta}_1)$
- Relationship between *F*-test and *t*-test for  $H_0: \beta_1 = 0$ :
  - $\triangleright$  Direct connection between test statistics:

$$F = t(\hat{\beta}_1)^2$$

 $\triangleright$  Importance: Connect

information about  $\hat{\beta}_1$  (captured by  $t(\hat{\beta}_1)$ ) with

information about the whole model (captured by F)

#### 1.5 Prediction

• Target (random variable):

$$y_* = \beta_0 + \beta_1 x_* + \varepsilon_*,$$

where  $x_*$  is explanatory variable value of interest

• Generic setting:

	response	known values of explanatory variables
	$\underline{y}$	<u>x</u>
	$y_1$	$x_1$
observed	$y_2$	$x_2$
(past) data	:	:
	$y_n$	$x_n$

Unobserved	$y_*$ (target)	$\leftarrow$	$x_*$
(future) data			

• Point predictor:

$$\hat{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_*$$

(Mnemonic: Set  $\beta_0 \to \hat{\beta}_0$ ,  $\beta_1 \to \hat{\beta}_1$ , and  $\varepsilon_* \to 0$ ; same trick as fitted values)

•  $100(1-\alpha)\%$  prediction interval:

point predictor  $\pm t$ -quantile  $\times$  st. error of prediction error  $= \hat{y}_* \pm t_{n-2,\alpha/2} \times \text{SE}(y_* - \hat{y}_*)$   $= (\hat{\beta}_0 + \hat{\beta}_1 x_*) \pm t_{n-2,\alpha/2} \sqrt{s^2 \left[1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}}\right]}$ 

(Suggestion: Remember this formula!)

• **Remarks on the structure of the prediction interval:** Two sources of uncertainty associated with prediction:

$$\widehat{\operatorname{Var}}(y_* - \hat{y}_*) = \underbrace{s^2}_{(1)} + \underbrace{s^2 \left[\frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}}\right]}_{(2)}.$$

- ① Variability of the random error  $\varepsilon_*$ : Reflected in the extra  $s^2$
- (2) Estimation of the true regression line at  $x_*$ :
  - $\triangleright \hat{\beta}_0$  and  $\hat{\beta}_1$  are only estimates of  $\beta_0$  and  $\beta_1$ , and are subject to sampling fluctuations
  - $\triangleright$  Variance of prediction error minimized when  $x_* = \bar{x}$  and increases quadratically as  $x_*$  moves away from  $\bar{x}$