Stat13-lecture 25 regression (continued, SE, t and chi-square)

- Simple linear regression model:
- $Y = \square_0 + \square_1 X + \square$
- Assumption : □is normal with mean 0 variance $\square^2$

  The fitted line is obtained by minimizing the sum of squared residuals; that is finding $\square_0$ and $\square_1$ so that

$(Y_1 - \square_0 - \square_1 X_1)^2 + \ldots. (Y_n - \square_0 - \square_1 X_n)^2$ is as small as possible

- This method is called least squares method

Least square line is the same as the regression line discussed before

- It follows that estimated slope β_1 can be computed by
- r [SD(Y)/SD(X)] = [cov(X,Y)/SD(X)SD(Y)][SD(Y)/SD(X)]
- =cov(X,Y)/VAR(X) (this is the same as equation for hat β_1 on page 518)
- The intercept β_0 is estimated by putting x=0 in the regression line; yielding equation on page 518
- Therefore, there is no need to memorize the equation for least square line; computationally it is advantageous to use cov(X,Y)/var(X) instead of r[SD(Y)/SD(X)]

Finding residuals and estimating the variance of □

- Residuals = differences between Y and the regression line (the fitted line)
- An unbiased estimate of $□^2$ is
- [sum of squared residuals]/ (n-2)
- Which divided by (n-2) ?
- Degree of freedom is n-2 because two parameters were estimated
- [sum of squared residuals]/$□^2$ follows a chi-square.

Hypothesis testing for slope

- Slope estimate $\hat{\beta}_1$ is random
- It follows a normal distribution with mean

equal to the true β_1 and the variance

equal to σ^2 / [n var(X)]

Because σ^2 is unknown, we have to estimate from the data ; the SE (standard error) of the slope estimate is equal to the squared root of the above

t-distribution

- Suppose an estimate hat $\hat{\square}$ is normal with
- variance c $\square^2$.
- Suppose $\square^2$ is estimated by s^2 which is related to a chi-squared distribution
- Then $(\hat{\square} - \square)/\sqrt{(c\ s^2)}$ follows a

t-distribution with the degrees of freedom equal to the chi-square degree freedom

An example

- Determining small quantities of calcium in presence of magnesium is a difficult problem of analytical chemists. One method involves use of alcohol as a solvent.
- The data below show the results when applying to 10 mixtures with known quantities of CaO. The second column gives
- Amount CaO recovered.
- Question of interest : test to see if intercept is 0 ; test to see if slope is 1.

X:CaO present	Y:CaO recovered	Fitted value	residual
4.0	3.7	3.751	-.051
8.0	7.8	7.73	.070
12.5	12.1	12.206	-.106
16.0	15.6	15.688	-.088
20.0	19.8	19.667	.133
25.0	24.5	24.641	-.141
31.0	31.1	30.609	.491
36.0	35.5	35.583	-.083
40.0	39.4	39.562	-.161
40.0	39.5	39.562	-.062

Least Squares Estimates:

	Estimate	Standard error
• Constant	-0.228090	(0.137840)
• Predictor	0.994757	(5.219485E-3)

- R Squared: 0.999780 ← Squared correlation
- Sigma hat: 0.206722 ← Estimate of SD(□)
- Number of cases: 10
- Degrees of freedom: 8

.22809/ .1378 =1.6547

(1 - 0.994757)/ 5.219485E-3 =
1.0045052337539044