ST5202: Applied Regression Analysis

Department of Statistics and Applied Probability
National University of Singapore

29-Jan-2018
Week 3

Announcement

- Assignment #1 due today
- Midterm scheduled on **12 March** at 7:00pm at **LT28**
- Make-up exam:
  - must make request by **26 February**. If you fail to make request by this date, no make-up exam will be available.
  - required to provide official supporting document (e.g., business trip, military service)

Week 3: Inference in Regression Analysis (Part 2)

- Review
- Inferences on β_1 (Continued)
  - Hypothesis testing
- Inferences Concerning β_0
- Interval Estimation of $E\{Y_h\}$
- Prediction of New Observation
- Confidence Band for Regression Line
- Analysis of Variance Approach to Regression Analysis
- General Linear Test Approach
- Descriptive Measures of Linear Association between X and Y
- Normal Correlation Model

Quick review: hypothesis testing

- Elements of a statistical test
  - Null hypothesis, H_0
  - Alternative hypothesis, H_a
  - Test statistic
  - Rejection region

Quick review: hypothesis testing

- Errors
  - Type I error: H_0 is rejected when H_0 is true
  - Type II error: H_0 is accepted when H_a is true

	H_0 is true	H_a is true
Accept H_0	Right Decision	Type II error
Reject H_0	Type I error	Right Decision

- p-value
  - The p-value, or attained significance level, is the smallest level of significance α at which the null hypothesis can be rejected from the observed data.

Review of Week 2:
Model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

- Y_i: value of the response variable of the i^{th} observation
- β_0, β_1: parameters
  β_1: slope, β_0: intercept
- ϵ_i are independent $N(0, \sigma^2)$
  Thus, $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

Review of Week 2:
Sampling distribution of $\frac{b_1 - \beta_1}{s\{b_1\}}$

$$\frac{(b_1 - \beta_1)}{s\{b_1\}} \sim t(n-2)$$

Hypothesis tests concerning β_1

- In many applications, the main interest is to investigate whether β_1 equals a fixed value, say β_{10}.
  (e.g., β_1=0 for $\beta_{10} = 0$, which indicates there is no linear association between X and Y)
- Two-sided test (for $\beta_{10} = 0$)

$$H_0 : \beta_1 = 0 \text{ vs. } H_a : \beta_1 \neq 0$$

- One-sided test (for $\beta_{10} = 0$)

$$\begin{aligned} H_0 &: \beta_1 \leq 0 \quad \text{vs.} \quad H_a : \beta_1 > 0 \\ \text{or } H_0 &: \beta_1 \geq 0 \quad \text{vs.} \quad H_a : \beta_1 < 0 \end{aligned}$$

Two-sided test: $H_0 : \beta_1 = \beta_{10}$ vs. $H_a : \beta_1 \neq \beta_{10}$

- Test statistic: $t^* = \frac{b_1 - \beta_{10}}{s\{b_1\}}$ $\left(s\{b_1\} = \sqrt{MSE / \sum_{i=1}^{n} (X_i - \bar{X})^2} \right)$
- If H_0 holds, then t^* is drawn from the sampling distribution centered at β_{10}, and

$$t^* = \frac{b_1 - \beta_{10}}{s\{b_1\}} \sim t(n-2)$$

- The decision rule:

$$\begin{aligned} &\text{If } |t^*| \leq t(1 - \alpha/2; n-2), \text{ conclude } H_0 \\ &\text{If } |t^*| > t(1 - \alpha/2; n-2), \text{ conclude } H_a \end{aligned}$$

- Note the relation with confidence interval

One-sided test: $H_0 : \beta_1 \leq \beta_{10}$ vs. $H_a : \beta_1 > \beta_{10}$

- Test statistic: $t^* = \frac{b_1 - \beta_{10}}{s\{b_1\}}$ $\left(s\{b_1\} = \sqrt{MSE / \sum_{i=1}^{n} (X_i - \bar{X})^2} \right)$
- The decision rule:

$$\text{If } t^* \leq t(1-\alpha; n-2), \text{ conclude } H_0$$
$$\text{If } t^* > t(1-\alpha; n-2), \text{ conclude } H_a$$

GPA vs. Entrance test score example

- $b_1 = 0.03883$, $s\{b_1\} = 0.01277$, and $t^* = \frac{0.03883-0}{0.01277} = 3.040$
- $t(1-0.05/2, 118) = 1.98027$, and $t(1-0.05, 118) = 1.65788$
- Two-sided test: $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ with $\alpha = 0.05$

$$\begin{aligned} |t^*| = |3.040| \quad &> \quad 1.98027 = t(1-0.05/2, 118) \\ \Longrightarrow \quad &\text{reject} \quad H_0 \end{aligned}$$

  - p-value: $P(|t(n-2)| > t^*) = 0.00292$

- One-sided test: $H_0 : \beta_1 \leq 0$ vs. $H_a : \beta_1 > 0$ with $\alpha = 0.05$

$$\begin{aligned} t^* = 3.040 \quad &> \quad 1.65788 = t(1-0.05, 118) \\ \Longrightarrow \quad &\text{reject} \quad H_0 \end{aligned}$$

  - p-value: $P(t(n-2) > t^*) = 0.001457$

Week 3: Inference in Regression Analysis (Part 2)
Hypothesis tests concerning β_1

GPA vs. Entrance test score example (continued)

R Console

```
lm(formula = Y ~ X, data = gpa.example)

Coefficients:
(Intercept)            X
    2.11405      0.03883

> summary(lm.gpa)

Call:
lm(formula = Y ~ X, data = gpa.example)

Residuals:
     Min       1Q   Median       3Q      Max
-2.74004 -0.33827  0.04062  0.44064  1.22737

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.11405    0.32089   6.588  1.3e-09 ***
X            0.03883    0.01277   3.040  0.00292 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6231 on 118 degrees of freedom
Multiple R-squared:  0.07262,    Adjusted R-squared:  0.06476
```

Inference on β_0:

The framework is the same as in case of β_1

- The sampling distribution of $\frac{b_0-\beta_0}{s\{b_0\}}$ is $t(n-2)$ where $s^2\{b_0\} = MSE[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X-\bar{X})^2}]$
- The $1-\alpha$ confidence interval for β_0 is

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\}$$

Considerations on inference on β_1 & β_0

- Normality assumption
  - The sampling distributions rely on the normality assumption on Y
  - If the probability distributions Y_i are not exactly normal but do not depart seriously, then the distribution of b_0 and b_1 will be approximately normal
  - Though Y_is are far from normal, for sufficiently large sample the distribution of b_0 and b_1 will be approximately normal (under general conditions)
- Spacing of the X levels:
  the variance of b_0 and b_1 (for fixed n and σ^2) strongly depends on the spacing of X due to the term $\sum(X_i - \bar{X})^2$

GPA vs. Entrance test score example

- $b_0 = 2.11405$, $s\{b_0\} = 0.32089$
- $t(1 - 0.05/2, 118) = 1.98027$
- 95% confidence interval

$$\begin{array}{ccc} 2.11405 & \pm & 1.98027 \cdot 0.32089 \\ (1.4786 & , & 2.7495) \end{array}$$

Interval Estimation of $E\{Y_h\}$

- X_h denotes the level of X for which we would like an estimate of the mean response
- The mean response when $X = X_h$ is denoted by

$$E\{Y_h\} = \beta_0 + \beta_1 X_h$$

- The point estimate of $E\{Y_h\}$ is

$$\hat{Y}_h = b_0 + b_1 X_h$$

Sampling distributions

- The sampling distribution is

$$\hat{Y}_h \sim N(E\{\hat{Y}_h\}, \sigma^2\{\hat{Y}_h\})$$

since $b_0 \sim N(\beta_0, Var\{b_0\})$ and $b_1 \sim N(\beta_1, Var\{b_1\})$

- What is the value of $E\{\hat{Y}_h\}$ and $\sigma^2\{\hat{Y}_h\}$?

Sampling distributions

- $E\{\hat{Y}_h\} = E\{b_0 + b_1 X_h\} = E\{b_0\} + E\{b_1\}X_h = \beta_0 + \beta_1 X_h = E\{Y_h\}$
- $Var\{\hat{Y}_h\} = \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)$ since

$$\begin{aligned}
Cov(\bar{Y}, b_1) &= Cov\left(\frac{1}{n}\sum Y_i, \sum k_i Y_i\right) \text{ where } k_i = \frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2} \\
&= \frac{1}{n}\sum k_i Var(Y_i) \\
&= \frac{\sigma^2}{n}\sum k_i = 0
\end{aligned}$$

$$\begin{aligned}
Var\{\hat{Y}_h\} &= Var\{\bar{Y} + b_1(X_h - \bar{X})\} \\
&= Var(\bar{Y}) + (X_h - \bar{X})^2 Var(b_1) + 2(X_h - \bar{X})Cov(\bar{Y}, b_1) \\
&= Var(\bar{Y}) + (X_h - \bar{X})^2 Var(b_1) \\
&= \frac{\sigma^2}{n} + \sigma^2 \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}
\end{aligned}$$

Sampling distributions

- $s^2\{\hat{Y}_h\} = MSE\left(\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right)$
- The sampling distribution of the studentized statistic is as follows:

$$\frac{\hat{Y}_h - E\{Y_h\}}{s\{\hat{Y}_h\}} \sim t(n-2)$$

Confidence interval for $E\{Y_h\}$

- Confidence interval:

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\hat{Y}_h\}$$

- From this, hypothesis test can be constructed in the usual manner

Comments

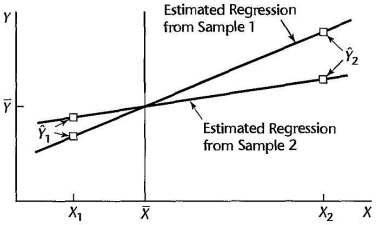


Figure: Effect on $\hat{Y}_h$ of variation in b_1 from sample to sample in two samples with same means $(\bar{X}, \bar{Y})$.

- The variance of the estimator for $E\{Y_h\}$ is smallest near the mean of X. Designing studies such that the mean of X is near X_h will improve inference precision
- When X_h is zero the variance of the estimator for $E\{Y_h\}$ reduces to the variance of the estimator b_0

GPA vs. Entrance test score example

- $\hat{Y}_h$ at $X = 27$: $2.11405 + 0.03883 \cdot 27 = 3.16238$
- $s\{\hat{Y}_h\}$ at $X = 27$: $0.6231 \cdot \sqrt{\frac{1}{120} + \frac{(27-24.725)^2}{2379.925}} = 0.063873$
- 95% confidence interval of $E[Y|X = 27]$:

$$\begin{array}{rcl} 3.16238 & \pm & 1.980272 \cdot 0.063873 \\ (3.035890 & , & 3.288873) \end{array}$$

Prediction of a new observation $Y_{h(new)}$

- $Y_{h(new)}$ denotes a new observation at given level X_h
- Inference on $E\{Y_h\}$ is making an inference on a population mean at given level X_h. On the other hand, $Y_{h(new)}$ is a single (future) observation
- $Y_{h(new)}$ is distributed around $E\{Y_h\}$

Properties

- $Y_{h(new)} - \hat{Y}_h \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})}{\sum(X_i - \bar{X})^2}\right)\right)$
  The normal distribution comes from the fact that
  $Y_{h(new)} \sim N\left(E\{Y_h\}, \sigma^2\right)$ and $\hat{Y}_h \sim N(E\{Y_h\}, Var\{\hat{Y}_h\})$
- Extra σ^2 in $Var(Y_{h(new)} - \hat{Y}_h)$ is from $\epsilon_{h(new)}$ where
  $Y_{h(new)} = \beta_0 + \beta_1 X_h + \epsilon_{h(new)}$

Properties

- $\frac{Y_{h(new)} - \hat{Y}_h}{s\{pred\}} \sim t(n-2)$
  - The numerator represents how far the new observation $Y_{h(new)}$ will deviate from the estimated mean $\hat{Y}_h$ based on the original n cases in the study
  - The numerator can be viewed as the prediction error
- $s^2\{pred\}$ represents an estimated variance of the numerator $Y_{h(new)} - \hat{Y}_h$
  - $s^2\{pred\} = MSE\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)$

Properties

- $Var\{pred\} = Var\{Y_{h(new)} - \hat{Y}_h\} = Var\{Y_{h(new)}\} + Var\{\hat{Y}_h\} = \sigma^2 + Var\{\hat{Y}_h\}$
- $Var\{pred\}$ has two components
  - The variance of the distribution of Y at $X = X_h$, namely σ^2
  - The variance of the sampling distrubtion of $\hat{Y}_h$, namely $Var\{\hat{Y}_h\}$
- An unbiased estimator of $\sigma^2\{pred\}$ is:

$$\begin{aligned} s^2\{pred\} &= MSE + s^2\{\hat{Y}_h\} \\ &= MSE\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right) \end{aligned}$$

Prediction limits

- From $\frac{Y_{h(new)}-\hat{Y}_h}{s\{pred\}} \sim t(n-2)$, the $1-\alpha$ prediction limits for a new observation $Y_{h(new)}$ is

$$\hat{Y}_h \pm t(1-\alpha/2, n-2)s\{pred\}$$

- Remark: prediction limit is different from confidence limit. We can make inference about an unknown (fixed) parameter (e.g., $E\{Y_h\}$), and construct confidence intervals of it. However, $Y_{h(new)}$ is not a parameter but a random value, about which we make predictions.

Prediction limits for mean of m new observations

- The $1-\alpha$ prediction limits for the mean of m new observations at given X_h :

$$\hat{Y}_h \pm t(1-\alpha/2, n-2)s\{predmean\}$$

- Here,

$$\begin{aligned} s^2\{predmean\} &= \frac{MSE}{m} + s^2\{\hat{Y}_h\} \\ &= MSE\left(\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right) \end{aligned}$$

GPA vs. Entrance test score example

- $\hat{Y}_h$ at $X = 27$: $2.11405 + 0.03883 \cdot 27 = 3.16238$
- $s\{pred\}$ at $X = 27$: $\sqrt{0.6231^2 + 0.063873^2} = 0.6263652$
- 95% prediction interval of Y_{new} at $X = 27$:

$$\begin{array}{rcl} 3.16238 & \pm & 1.980272 \cdot 0.6263652 \\ (1.921958 & , & 4.402805) \end{array}$$

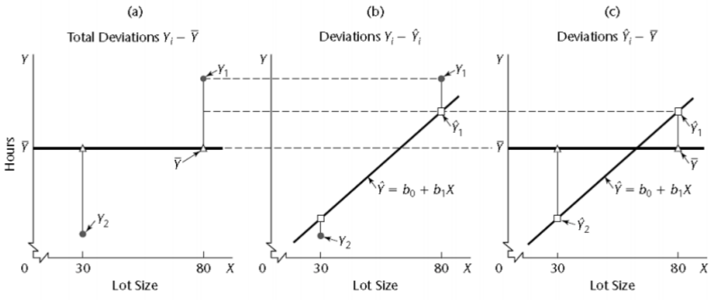
GPA vs. Entrance test score example

R Console

```
>
> ## Constructing confidence interval for E[Y_h] at X=27 and X=32
> newobs = data.frame(Y=c(NA, NA), X=c(27, 32))
> predict(lm.gpa, new = newobs, interval = "confidence")
       fit      lwr      upr
1 3.162382 3.035890 3.288873
2 3.356517 3.140763 3.572272
>
> ## Constructing prediction interval for E[Y_h] at X=27 and X=32
> predict(lm.gpa, new = newobs, interval = "prediction")
       fit      lwr      upr
1 3.162382 1.921958 4.402805
2 3.356517 2.103840 4.609195
>
```

Analysis of Variance (ANOVA) approach

Partitioning of total sum of squares



$$Y_i - \bar{Y} = \left(Y_i - \hat{Y}_i\right) + \left(\hat{Y}_i - \bar{Y}\right)$$

Partitioning of total sum of squares

- Total Sum of Squares (SSTO): $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$
  - The measure of total variation
  - If all Y_i's are the same, then $SSTO = 0$
- Error Sum of Squares (SSE): $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$
  - The measure of variations of the Y_i's that is still present when the predictor variable X is taken into account
- Regression sum of squares (SSR): $\sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2$
  - The measure of variation of the Y_i's associated with the regression line
  - $\sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2 = b_1^2 \sum_{i=1}^{N}(X_i - \bar{X})^2$

Partitioning of total sum of squares

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$$

- $Y_i - \hat{Y}_i$: the deviation of the observation Y_i around the fitted regression line
- $\hat{Y}_i - \bar{Y}$: the deviation of the fitted value $\hat{Y}_i$ around the mean $\bar{Y}$

Partitioning of total sum of squares

$$\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2$$

or equivalently

$$\text{SSTO} = \text{SSE} + \text{SSR}$$

Partitioning of total sum of squares: proof

$$\sum(Y_i - \bar{Y})^2 \;=\; \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2 + 2\sum(Y_i - \hat{Y}_i)\cdot(\hat{Y}_i - \bar{Y})$$

Here, $\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \sum \hat{Y}_i(Y_i - \hat{Y}_i) + \sum \bar{Y}(Y_i - \hat{Y}_i) = 0$:

- The first term: $\sum \hat{Y}_i(Y_i - \hat{Y}_i) = \sum \hat{Y}_i e_i = 0$ by (1.20)
- The second term: $\sum \bar{Y}(\hat{Y}_i - \bar{Y}) = \bar{Y}\sum e_i = 0$

Breakdown of degrees of freedom

- SSTO: n-1 degrees of freedom
  1 linear constraint due to the calculation and inclusion of the mean
- SSE: n-2 degrees of freedom
  2 linear constraints due to estimating β_0 and β_1
- SSR: 1 degree of freedom
  Two degrees of freedom in regression parameters, and one is lost due to linear constraint
- $n-1=(n-2)+(1)$

Mean Squares

A sum of squares divided by its associated degrees of freedom is called a mean square

- The regression mean square:

$$MSR = \frac{SSR}{1}$$

- The mean square error:

$$MSE = \frac{SSE}{n-2}$$

Expected mean squares

$$
\begin{aligned}
E\{MSE\} &= \sigma^2 \\
E\{MSR\} &= \sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2
\end{aligned}
$$

- The mean of the sampling distribution of MSE is σ^2 whether or not X and Y are linearly correlated
- The mean of the sampling distribution of MSR is σ^2 when $\beta_1 = 0$. Hence if $\beta_1 = 0$ holds, MSR and MSE will tend to have the same order of magnitude

Expected mean squares

- $E\{MSE\} = \sigma^2$. We have seen it in previous slides
-

$$
\begin{aligned}
E\{MSR\} &= E\{SSR\} = E\{b_1^2 \sum (X_i - \bar{X})^2\} \\
&= \sum (X_i - \bar{X})^2 E\{b_1^2\} \\
&= \sum (X_i - \bar{X})^2 \left(\frac{\sigma^2}{\sum (X_i - \bar{X})^2} + \beta_1^2 \right) \\
&= \sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2
\end{aligned}
$$

Here, we have

$$
\begin{aligned}
E\{b_1^2\} &= Var\{b_1\} + E\{b_1\}^2 \\
&= \left(\frac{\sigma^2}{\sum (X_i - \bar{X})^2} \right) + (\beta_1)^2
\end{aligned}
$$

F Test of $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$

- Hypothesis: $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$
- Test statistic: $F^* = \frac{MSR}{MSE}$
  - Note the different form from $\frac{b_1 - 0}{s\{b_1\}}$ of which the sampling distribution is $t(n-2)$
- Sampling distribution of F^*:

$$F^* \sim F(1, n-2) \text{ when } H_0 : \beta_1 = 0 \text{ holds}$$

Sampling distribution of F^*

- The sampling distribution of F^* when $H_0 : \beta_1 = 0$ holds can be derived from Chchran's theorem
- Cochran's theorem:
  if all n observations Y_i come from the same normal distribution with mean μ and variance σ^2, and $SSTO$ is decomposed into k sums of squares SS_r, each with degrees of freedom df_r, then the SS_r/σ^2 terms are independent χ^2 variable with df_r degrees of freedom if $\sum_{r=1}^{k} df_r = n - 1$
  - $SSTO(df = n-1) = SSE(df = n-2) + SSR(df = 1)$ with $n - 1 = (n-2) + (1)$
  - If $\beta_1 = 0$, then Y_i have the same mean $\mu = \beta_0$ and the same variance σ^2
  - Therfore, from Cochran's theorem, if $\beta_1 = 0$ we have SSE/σ^2 and SSR/σ^2 are independent χ^2 variables with degrees of freedom n-2 and 1 respectively

Sampling distribution of F^*

- For two independent random variables W_m and W_n where $W_m \sim \chi^2(m)$ and $W_n \sim \chi^2(n)$,

$$\frac{W_m/m}{W_n/n} \sim F(m, n)$$

- We have $SSR/\sigma^2 \sim \chi^2(1)$, $SSE/\sigma^2 \sim \chi^2(n-2)$, and $SSR/\sigma^2 \perp SSE/\sigma^2$ when $\beta_1 = 0$. Therefore

$$F^* = \frac{SSR/\sigma^2}{1} \Big/ \frac{SSE/\sigma^2}{n-2} \sim F(1, n-2) \text{ when } H_0 : \beta_1 = 0 \text{ holds}$$

Decision rule

$$\text{If } F^* \le F(1-\alpha; 1, n-2), \text{ conclude } H_0$$
$$\text{If } F^* > F(1-\alpha; 1, n-2), \text{ conclude } H_a$$

- $F(1-\alpha; 1, n-2)$ denotes the $(1-\alpha)100$ percentile of the $F(1, n-2)$ distribution
- Controls the risk of Type I error to be α
- The test is upper-tail

Equivalence of F Test and two-sided t Test for $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$

- We have

$$F^* = \frac{MSR}{MSE} = \frac{b_1^2 \sum(X_i - \bar{X})^2}{MSE} = \frac{b_1^2}{MSE/\sum(X_i - \bar{X})^2} = \frac{b_1^2}{s\{b_1\}^2} = (t^*)^2$$

$(s^2\{b_1\} = MSE/\sum(X_i - \bar{X})^2)$

- Also, $t(m)^2 \sim (\frac{z}{\sqrt{W_m/m}})^2 \sim \frac{W_1/1}{W_m/m} \sim F(1, m)$ where $z \sim N(0, 1)$, $W_1 \sim \chi^2(1)$, $W_m \sim \chi^2(m)$, and z, W_1, W_m are all independent. This leads to

$$[t(1 - \alpha/2; n - 2)]^2 = F(1 - \alpha; 1, n - 2)$$

Equivalence of F Test and two-sided t Test for $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ (continued)

- Thus, for any α

  Accept H_0: $\{F^* \leq F(1-\alpha; 1, n-2)\}$ equiv. to $\{|t^*| \leq t(1-\alpha/2; n-2)\}$

  Accept H_a: $\{F^* > F(1-\alpha; 1, n-2)\}$ equiv. to $\{|t^*| > t(1-\alpha/2; n-2)\}$

ANOVA table

Source	SS	df	MS	F	p-value(s)
Regression	SSR	2-1	MSR	$F^* = \frac{MSR}{MSE}$	$P(F(1, n-2) \geq f^*)$
Error	SSE	$n-2$	MSE		
Total	SSTO	$n-1$			

- One of the important role of the ANOVA table above is to test $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$

GPA vs. Entrance exam score example

Source	SS	df	MS	F	p-value
Regression	3.59	1	3.588	9.25	0.00292
Error	45.82	118	0.388		
Total	49.41	119			

- $F^* = \frac{3.588/1}{45.82/118} = \frac{3.588}{0.388} = 9.25$
- $F^* = 9.25 > 3.921478 = F(1 - 0.05; 1, n - 2)$
  Therefore, we reject $H_0 : \beta_1 = 0$ with $\alpha = 0.05$
- Also, $(t^*)^2 = 3.04^2 = 9.24$

Week 3: Inference in Regression Analysis (Part 2)
ANOVA

GPA vs. Entrance exam score example

R Console

```
> aov.gpa = aov(lm.gpa)
> summary(aov.gpa)
             Df Sum Sq Mean Sq F value  Pr(>F)
X             1   3.59   3.588    9.24 0.00292 **
Residuals   118  45.82   0.388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

General linear test approach

- Three steps:
  1. Full model
  2. Reduced model
  3. Test statistic

- Testing $\beta_1 = 0$ vs $\beta_1 \neq 0$ is a kind of general linear test approach

Full model

- Full or unrestricted model is a model that is considered to be appropriate for the data
  - Full model for the simple linear regression is the usual normal error regression model:
    $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- Error sum of squares of the full model ($SSE(F)$) measures the variability of the Y_i observations around the fitted regression line from the full model
  - For simple linear regression,
    $SSE(F) = \sum[Y_i - \hat{Y}_i]^2 = \sum[Y_i - (b_0 + b_1 X_i)]^2 = SSE$

Reduced model

- The model when H_0 holds is called the reduced or restricted model
  - For testing $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$, the reduced model is

$$Y_i = \beta_0 + \epsilon_i$$

- The error sum of squares ($SSE(R)$) is the variability of the observation Y_i around the fitted regression line from the reduced model
  - For the reduced model under $H_0 : \beta_1 = 0$, the LS or maximum likelihood estimator of β_0 by is $b_0 = \bar{Y}$.
    Thus, $SSE(R) = \sum(Y_i - b_0)^2 = \sum(Y_i - \bar{Y})^2 = SSTO$

Test statistic

- It always holds that $SSE(F) \leq SSE(R)$ since the more parameters in the model, the better the one can fit the data
- IDEA: if $SSE(F)$ is not much less that $SSE(R)$, then it implies the full model does not explain the data much better than the reduced model and the data is in favor of H_0
  - a small difference $SSE(R) - SSE(F)$ supports H_0
  - a large difference $SSE(R) - SSE(F)$ supports H_a

Test statistic

- The test statistic
$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \sim F(df_R - df_F, df_F) \text{ when } H_0 \text{ holds}$$
where df_R and df_F are the degrees of freedom associated with the reduced model and the full model respectively
- The decision rule:
$$\text{If } F^* \leq F(1-\alpha; df_R - df_F, df_F), \text{ conclude } H_0$$
$$\text{If } F^* > F(1-\alpha; df_R - df_F, df_F), \text{ conclude } H_a$$
- For simple linear regression with $H_0 : \beta_1 = 0$,
$$SSE(R) = SSTO \qquad SSE(F) = SSE$$
$$df_R = n-1 \qquad df_F = n-2$$
$$F^* = \frac{SSTO - SSE}{(n-1)-(n-2)} \div \frac{SSE}{n-2} = \frac{SSR}{1} \div \frac{SSE}{n-2} = \frac{MSR}{MSE}$$

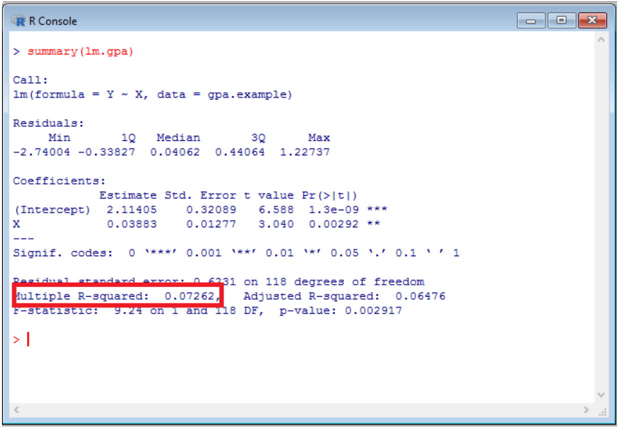
Descriptive measures of linear association between X and Y: Coefficient of Determination

- The coefficient of determination
  - SSTO: a measure of uncertainty of Y when X is not taken into account
  - SSE: a meaure of uncertainty of Y when X is taken into account
  - Coefficient of determination R^2:
    $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$
    reduction of uncertainty due to considering X
  - $0 \leq R^2 \leq 1$

Descriptive measures of linear association between X and Y

- $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} = \frac{[\sum(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}$
- Correlation coefficient: $r = \pm\sqrt{R^2} = \sqrt{\frac{[\sum(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$
  - if $b_1 > 0$, then $r = \sqrt{R^2}$
  - if $b_1 < 0$, then $r = -\sqrt{R^2}$
  - $-1 \leq r \leq 1$

Week 3: Inference in Regression Analysis (Part 2)
Descriptive Measures of Linear Association



R Console

```
> summary(lm.gpa)

Call:
lm(formula = Y ~ X, data = gpa.example)

Residuals:
     Min       1Q   Median       3Q      Max
-2.74004 -0.33827  0.04062  0.44064  1.22737

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.11405    0.32089   6.588  1.3e-09 ***
X            0.03883    0.01277   3.040  0.00292 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6231 on 118 degrees of freedom
Multiple R-squared:  0.07262,    Adjusted R-squared:  0.06476
F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917

>
```

Comments on R^2

- R^2 describes only relative reduction of variation via regression model, and does not indicate predictive power of the model
- R^2 only captures linear relationship between Y and X
- As R^2 cannot capture nonlinear relationship, nonlinear relationship may coexist with either high or low R^2
- High R^2 does not necessarily indicate strong linear relationship between X and Y
- Low R^2 does not necessarily indicate no relationship between Y and X

Normal correlation models

- Distinction between regression models and correlation models
  - Regression models: X values are fixed constants
  - Correlation models: both X and Y are random variables
- In some cases, correlation models are more suitable than regression models
  - Relationship between sales of gasoline and sales of auxiliary products
  - Relationship between blood pressure and weight

Bivariate normal

- Y_1 and Y_2 are jointly normally distributed if the joint distribution has the density of the bivariate normal distribution:

$$
\begin{aligned}
f(Y_1, Y_2) \quad = \quad & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left(-\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 \right.\right. \\
& \left.\left. + \left(\frac{Y_2-\mu_2}{\sigma_2}\right)^2 \right] - 2\rho_{12}\left(\frac{Y_1-\mu_1}{\sigma_1}\right)\left(\frac{Y_2-\mu_2}{\sigma_2}\right) \right)
\end{aligned}
$$

Bivariate normal: parameters

- Parameters
  - μ_1, μ_2: means of Y_1 and Y_2 respectively
  - σ_1, σ_2: standard deviations of Y_1 and Y_2 respectively
  - ρ_{12}: coefficient of correlation between the random variables Y_1 and Y_2

$$\rho_{12} = \frac{E\{(Y_1 - \mu_1)(Y_2 - \mu_2)\}}{\sqrt{Var\{Y_1\}Var\{Y_2\}}}$$

- Properties
  - $-1 \leq \rho_{12} \leq 1$
  - If $Y_1 \perp Y_2$ then $\rho_{12} = 0$
  - If Y_1 and Y_2 are positively correlated, then $\rho_{12} > 0$
  - If Y_1 and Y_2 are negatively correlated, then $\rho_{12} < 0$

Bivariate normal: conditional inference

- Conditional probability distribution of Y_1 given Y_2
  - $f(Y_1|Y_2) = \frac{f(Y_1,Y_2)}{f_2(Y_2)} = \frac{1}{\sqrt{2\pi}\sigma_{1|2}} \exp\left[-\frac{1}{2}\left(\frac{Y_1-\alpha_{1|2}-\beta_{12}Y_2}{\sigma_{1|2}}\right)^2\right]$

    where

$$\begin{aligned}\alpha_{1|2} &= \mu_1 - \mu_2\rho_{12}\frac{\sigma_1}{\sigma_2}\\ \beta_{12} &= \rho_{12}\frac{\sigma_1}{\sigma_2}\\ \sigma^2_{1|2} &= \sigma^2_1(1-\rho^2_{12})\end{aligned}$$

  - Thus, $Y_1|Y_2 \sim N(\alpha_{1|2} + \beta_{12}Y_2, \sigma^2_{1|2})$
  - $\alpha_{1|2}$ is the intercept of the line regression of Y_1 on Y_2
  - β_{12} is the slope of this line

Bivariate normal: conditional inference

- In the same manner, conditional probability distribution of Y_2 given Y_1 is
  - $f(Y_2|Y_1) = \frac{1}{\sqrt{2\pi}\sigma_{2|1}} \exp\left[-\frac{1}{2}\left(\frac{Y_1-\alpha_{2|1}-\beta_{21}Y_2}{\sigma_{2|1}}\right)^2\right]$
    where

$$\begin{aligned}
\alpha_{2|1} &= \mu_2 - \mu_1\rho_{12}\frac{\sigma_2}{\sigma_1}\\
\beta_{21} &= \rho_{12}\frac{\sigma_2}{\sigma_1}\\
\sigma_{2|1}^2 &= \sigma_2^2(1-\rho_{12}^2)
\end{aligned}$$

  - $Y_2|Y_1 \sim N(\alpha_{2|1} + \beta_{21}Y_1, \sigma_{2|1}^2)$
  - $\alpha_{2|1}$ is the intercept of the line regression of Y_2 on Y_1
  - β_{21} is the slope of this line

Important characteristics of conditional distributions

- The conditional probability distribution of Y_1 for any given value of Y_2 is normal
- The means of the conditional probability distributions of Y_1 fall on a straight line with respect to Y_2, and hence are a linear function of Y_2:

$$E\{Y_1|Y_2\} = \alpha_{1|2} + \beta_{12}Y_2$$

- All conditional probability distribution of Y_1 have the same standard deviation $\sigma_{1|2}$ regardless of the given value Y_2

Equivalence to normal error regression model

- For a bivariate normal random sample (Y_1, Y_2), the normal error regression model is applicable for conditional inference about Y_1 given Y_2:
  - The Y_1 observations are independent
  - The observations Y_1 given Y_2 are normally distributed with mean $E\{Y_1|Y_2\} = \alpha_{1|2} + \beta_{12}Y_2$ and constant variance $\sigma^2_{1|2}$

Inference on correlation coefficient

- Point estimator of ρ_{12}:

$$r_{12} = \frac{\sum(Y_{i1} - \bar{Y}_1)(Y_{i2} - \bar{Y}_2)}{\left[\sum(Y_{i1} - \bar{Y}_1)^2 \sum(Y_{i2} - \bar{Y}_2)^2\right]^{1/2}}$$

- Hypothesis

$$\begin{array}{lclcl} H_0 : \rho_{12} = 0 & \text{equiv. to} & H_0 : \beta_{12} = 0 & \text{equiv. to} & H_a : \beta_{21} = 0 \\ H_a : \rho_{12} \neq 0 & & H_a : \beta_{12} \neq 0 & & H_a : \beta_{21} \neq 0 \end{array}$$

- Test statistic: $t^* = \frac{r_{12}\sqrt{n-2}}{\sqrt{1-r_{12}^2}}$
- Decision rule

$$\text{If } |t^*| \leq t(1-\alpha/2; n-2), \text{ conclude } H_0$$
$$\text{If } |t^*| > t(1-\alpha/2; n-2), \text{ conclude } H_a$$

Interval estimation of ρ_{12}

- When $\rho_{12} \neq 0$, the sampling distribution of r_{12} complicated. Thus we use the *Fisher z transformation* :

$$z' = \frac{1}{2} \log_e \left(\frac{1 + r_{12}}{1 - r_{12}} \right)$$

- When n is large, the distribution of z' is approximately normal with mean and variance as follows:

$$\begin{aligned} E\{z'\} &= \zeta = \frac{1}{2} \log_e \left(\frac{1 + \rho_{12}}{1 - \rho_{12}} \right) \\ Var\{z'\} &= \frac{1}{n - 3} \end{aligned}$$

- Approximate $1 - \alpha$ confidence limits for ζ are

$$z' \pm z(1 - \alpha/2)\sigma\{z'\}$$

- The $1 - \alpha$ confidence limits for ρ_{12} are then obtained by transforming the limits on ζ utilizing the Fisher z transformation relation

Reading: entire Chapter 2