An Overview

- Inferences about β_1.
  1. Sampling distribution of $\hat{\beta}_1$.
  2. Sampling distribution of $\{\hat{\beta}_1 - \beta_1\}/\sqrt{\hat{\text{var}}\{\hat{\beta}_1\}}$.
  3. Confidence interval for β_1.
  4. Hypothesis testing.
- Inferences about β_0.
- Estimation and prediction (with respect to some x_0).
- ANOVA approach.
- Coefficient of determination: R^2.

Inferences about β_1

- Recall simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad \epsilon_i \sim \text{iid } N(0, \sigma^2),$$

  for $i = 1, \ldots, n$.

- Recall that LS and ML estimate of β_1 is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

- As we will show, $\hat{\beta}_1$ is normal with

$$E(\hat{\beta}_1) = \beta_1 \quad \text{and} \quad Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Preliminary Results Concerning $\hat{\beta}_1$

- Note that

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \sum_{i=1}^n \frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} Y_i\end{aligned}$$

- Express $\hat{\beta}_1$ as a linear combination of $\{Y_i\}$:

  $\hat{\beta}_1 =$

  where

Preliminary Results Concerning $\hat{\beta}_1$

Now we compute $\sum_{i=1}^{n} k_i$, $\sum_{i=1}^{n} k_i X_i$, and $\sum_{i=1}^{n} k_i^2$.

Sampling Distribution of $\hat{\beta}_1$

- $\hat{\beta}_1$ follows a normal distribution. Why?
- Mean of $\hat{\beta}_1$, $E(\hat{\beta}_1)$ is

- Variance of $\hat{\beta}_1$, $Var(\hat{\beta}_1)$ is

Completing Proof of Gauss-Markov Theorem

Theorem: Under the simple linear regression model with the residuals being mean zero, constant variance (but not necessarily normal), $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are **BLUE** (**Best Linear Unbiased Estimators**) because they have minimum variance among all linear unbiased estimators.

Proof:

- We have already shown that the estimators are linear since
- We have already shown that $\widehat{\beta}_1$ is unbiased, i.e., $E(\widehat{\beta}_1) = \beta_1$. Now
- It remains to show the minimum variance among all linear estimators.

$Var(\hat{\beta}_1)$ is minimal among unbiased linear estimators

- Let an arbitrary linear unbiased estimator be of the form $\tilde{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$ where c_i are constants that satisfy

- Note that Var$\{\tilde{\beta}_1\}$ is

$Var(\hat{\beta}_1)$ is minimal among unbiased linear estimators

- Also note that $\sum_{i=1}^{n} k_i d_i = 0$. Why? Show it below.

- Thus

Estimate of $Var(\hat{\beta}_1)$

- Recall that

$$\text{Var}\{\hat{\beta}_1\} = \sigma^2 \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 \right)^{-1}$$

- The estimated variance is

$$\hat{\text{Var}}\{\hat{\beta}_1\} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2/(n-2)}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{MSE}{\sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

- We will show

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\text{Var}}\{\hat{\beta}_1\}}} \sim t_{n-2}.$$

Sampling Distribution of $\frac{\hat{\beta}_1-\beta_1}{\sqrt{\hat{\text{Var}}\{\hat{\beta}_1\}}}$

- Definition: If $Z_i \sim$ i.i.d. $N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, k$ then $\sum_{i=1}^{k} \left(\frac{Z_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2_k$ (Chi-square distribution with k degrees of freedom).
- Definition (KNNL A.44): A t random variable with ν degrees of freedom results from the expression

$$t_\nu = \frac{z}{\sqrt{q_\nu/\nu}}$$

where z and q_ν are independent standard normal and χ^2_ν random variables, respectively.

Sampling Distribution of $\frac{\hat{\beta}_1-\beta_1}{\sqrt{\hat{\text{Var}}\{\hat{\beta}_1\}}}$

- Note that

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\text{Var}\{\hat{\beta}_1\}}}=\frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n(X_i-\bar{X})^2}}}\sim N(0,1).$$

- Also

$$\frac{\hat{\text{Var}}\{\hat{\beta}_1\}}{\text{Var}\{\hat{\beta}_1\}}=\frac{\frac{MSE}{\sum_{i=1}^n(X_i-\bar{X})^2}}{\frac{\sigma^2}{\sum_{i=1}^n(X_i-\bar{X})^2}}=\frac{MSE}{\sigma^2}=\frac{\text{SSE}}{(n-2)\sigma^2}\sim\frac{\chi^2_{n-2}}{n-2}.$$

- Thus,

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\hat{\text{Var}}\left\{\hat{\beta}_1\right\}}}=\frac{\frac{\hat{\beta}_1-\beta_1}{\sqrt{\text{Var}\left\{\hat{\beta}_1\right\}}}}{\sqrt{\frac{\hat{\text{Var}}\{\hat{\beta}_1\}}{\text{Var}\{\hat{\beta}_1\}}}}\sim t_{n-2}$$

Sampling Distribution of $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\text{Var}}\{\hat{\beta}_1\}}}$

- The last conclusion on the previous slide only holds if SSE/σ^2 is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.
- This is given in Theorem (2.11) of KNNL for the simple linear regression model and is proven in general in STAT640.

Confidence Interval for β_1

- Denote $\sqrt{\hat{\text{Var}}\left\{\hat{\beta}_1\right\}} = s\{\hat{\beta}_1\}$. Recall that

$$\frac{\hat{\beta}_1 - \beta_1}{s\{\hat{\beta}_1\}} \sim t_{n-2}.$$

- The $(1-\alpha)$ CI for β_1 is

$$\hat{\beta}_1 \pm t(1-\alpha/2; n-2)s\{\hat{\beta}_1\}$$

  where $t(1-\alpha/2; n-2)$ is the $(1-\alpha/2)100$th percentile of t_{n-2}.
- See KNNL pp.1317–1318 Table B.2. (Or use $\pm$ `qt(.025,df)`)
- Interpretation of CI. For example, a 95% CI for β_1 is (–,–).
  - If we repeated the study 100 times and created 100 CI's for β_1, we would expect that 95 of these intervals would include the true value of β_1.
  - The method used to construct this interval has a 5% error rate.

Confidence Interval for β_1

- In the advertising example, recall that

$$\sum_{i=1}^{n} X_i = 24.40, \sum_{i=1}^{n} X_i^2 = 107.42, \sum_{i=1}^{n} X_i Y_i = 154.07$$
$$\sum_{i=1}^{n} Y_i = 35.50, \sum_{i=1}^{n} Y_i^2 = 222.03.$$

- Thus

$$\begin{aligned} \hat{\beta}_1 &= \frac{154.07 - 24.40 \times 35.50/7}{107.42 - 24.4 \times 24.40/7} = \frac{30.33}{22.37} = 1.356 \\ \hat{\beta}_0 &= 35.50/7 - 1.356 \times 24.40/7 = 0.345 \end{aligned}$$

- Compute MSE

$$MSE = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2} = \frac{0.86}{5} = 0.172$$

Confidence Interval for β_1

- Compute standard deviation estimate

$$s\{\hat{\beta}_1\} = \sqrt{\frac{MSE}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \sqrt{\frac{0.172}{22.37}} = 0.0877$$

- For a 95% CI, $\alpha = 0.05$ and

$$t(1 - \alpha/2; n - 2) = t(0.975; 5) = 2.571$$

- The 95% CI for β_1 is

$$\begin{aligned} \hat{\beta}_1 \pm t(1 - \alpha/2; n - 2)s\{\hat{\beta}_1\} &= 1.356 \pm 2.571 \times 0.0877 \\ &= 1.356 \pm 0.225 \\ &= (1.13, 1.58). \end{aligned}$$

Review of Hypothesis Testing

- Recall two types of errors.
  - Type I: Reject H_0 when H_0 is true.
  - Type II: Fail to reject H_0 when H_0 is false.
- Level of significance $\alpha = P(\text{Type I error})$.
- Power $= 1 - P(\text{Type II error}) = 1 - \beta$.
- Recall p-value.
  - A p-value is the probability of observing a sample outcome as extreme or more extreme than the observed outcome under the assumption that H_0 is true.
  - Small p-value provides evidence against H_0.
  - It is misleading to say that p-value $= 0$. Use p-value ≤ 0.0001.
- When we choose α, we control P[type I error] but it will affect β, too. We can't choose both α and β (without manipulating n), so we choose to control α (more important).

Review of Hypothesis Testing

- 1-sided versus 2-sided test.
- CI versus hypothesis testing: For the t-test given above, we could state the conclusion of an α-level test in terms of a $(1-\alpha)100\%$ CI. If 0 is contained in the $(1-\alpha)100\%$ CI, then we fail to reject H_0.
- When writing up a hypothesis test for this class, always include
  - Hypothesis in statistical and practical terms.
  - Test statistic.
  - Decision rule and p-value.
  - Conclusion in the context of a problem. Use wording such as "reject H_0" or "fail to reject H_0". Do not use "accept H_0".

Hypothesis Testing for β_1

- A test of interest (why?) is:

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

- Recall that

$$\frac{\hat{\beta}_1 - \beta_1}{s\{\hat{\beta}_1\}} \sim t_{n-2}$$

- Thus an α-level test is based on the test statistic

$$t^* = \frac{\hat{\beta}_1 - 0}{s\{\hat{\beta}_1\}}.$$

- Decision rule: If $|t^*| > t(1 - \alpha/2; n - 2)$, reject H_0; otherwise do not reject H_0.
- p-value $= 2 \times P(T > |t^*|)$ where $T \sim t_{n-2}$.

Hypothesis Testing for β_1

- Revisit the advertising example and test

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

- The test statistic is

$$t^* = \frac{\hat{\beta}_1}{s\{\hat{\beta}_1\}} = \frac{1.356}{0.0877} = 15.46.$$

- Compared with t_5, the p-value is

$$2 \times P(t_5 > 15.46) < 0.0001.$$

- Thus reject H_0 and there is strong evidence that there is a positive line relationship between advertising expenditure and sales.

Hypothesis Testing for β_1

- In general, for testing $H_0 : \beta_1 = \beta_h$ versus $H_a : \beta_1 \neq \beta_h$, use the test statistic

$$t^* = \frac{\hat{\beta}_1 - \beta_h}{s\{\hat{\beta}_1\}}.$$

  and proceed as before.

- For testing $H_0 : \beta_1 \leq \beta_h$ versus $H_a : \beta_1 > \beta_h$
  - Decision rule: If $t^* > t(1-\alpha; n-2)$, reject H_0; otherwise do not reject H_0.
  - p-value $= P(T > t^*)$ where $T \sim t_{n-2}$
- For testing $H_0 : \beta_1 \geq \beta_h$ versus $H_a : \beta_1 < \beta_h$
  - Decision rule: If $t^* < t(\alpha; n-2)$, reject H_0; otherwise do not reject H_0.
  - p-value $= P(T < t^*)$ where $T \sim t_{n-2}$

Inferences about β_0

- Recall that

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

- It can be shown that $\hat{\beta}_0$ is normal with

$$E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad \sigma^2\{\hat{\beta}_0\} = Var(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right].$$

(Left as HW)

- Estimated variance is

$$s^2\{\hat{\beta}_0\} = MSE \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right].$$

Inferences about β_0

- It can be shown that the sampling distribution is

$$\frac{\hat{\beta}_0 - \beta_0}{s\{\hat{\beta}_0\}} \sim t_{n-2}.$$

- CIs and hypothesis tests for β_0 follow as those for β_1.
- Note the case of $\beta_0 = 0$.
- In practice, never drop β_0 from the model unless there is a scientific reason.
- However, rarely is one interested in the actual value of $\widehat{\beta_0}$

Types of Prediction and Estimation

- Estimate the mean response $E(Y_h)$ for a given level of $X = X_h$.
- Predict a new observation $Y_{h(\text{new})}$ for a given level of $X = X_h$.
- Predict the mean of m new observations all at a given level of X_h.
- Estimate confidence band for regression line for several (or all) X_h's.

Data Example: muscle mass (HW 1 (7))

Recall that $Y =$ muscle mass and $X =$ age with the fitted regression line

$$\hat{Y} = 156.35 - 1.19X$$

- What is the population mean measure of muscle mass for a 55-year-old person?
- What should we predict for the muscle mass for a 55-year-old person randomly selected from the population?
- In both cases, the estimate is

$$\hat{Y} = 156.35 - 1.19 \times 55 = 90.9$$

but uncertainty is larger in the second case.

Estimation of $E(Y_h)$

- Let X_h be the level of X for which we want to estimate the mean response.
- X_h could be observed or not, but should be within the range of $\{X_i\}$.
- $E(Y_h) =$ the mean response at X_h.
- The estimate of $E(Y_h)$ is

Derivation of $Var(\hat{Y}_h)$

Three results are used in the derivation.

$$\hat{Y}_h =$$

$$Var(a_1Y_1 + a_2Y_2) =$$

$$Cov(\bar{Y}, \hat{\beta}_1) =$$

Derivation of $Var(\hat{Y}_h)$

Thus, $\sigma^2\{\hat{Y}_h\}$ is

Inference for $E(Y_h)$

- Variance of $\hat{Y}_h$ is

$$\sigma^2\{\hat{Y}_h\} = Var(\hat{Y}_h) = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right].$$

- Estimated variance is

$$s^2\{\hat{Y}_h\} = MSE\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right].$$

- Note that

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

- The $(1-\alpha)$ CI for $E(Y_h)$ is

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

Inference for $E(Y_h)$

- In the advertising example, suppose $X_h = 6$.
- The estimate of the mean sales at $X_h = 6$ is

$$\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = 0.345 + 1.356 \times 6 = 8.48.$$

- The estimated standard deviation is

$$\begin{aligned} s\{\hat{Y}_h\} &= \sqrt{MSE\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right]} \\ &= \sqrt{0.172} \times \sqrt{\frac{1}{7} + \frac{(6 - 3.486)^2}{22.37}} = 0.271. \end{aligned}$$

- The 95% CI for the mean sales at $X_h = 6$ is

$$\begin{aligned} \hat{Y}_h \pm t(1 - \alpha/2; n-2) s\{\hat{Y}_h\} &= 8.48 \pm 2.571 \times 0.271 \\ &= 8.48 \pm 0.70 = (7.78, 9.18). \end{aligned}$$

Inference for $Y_{h(\text{new})}$

- X_h = the "new" value of X.
  - In the previous case, X_h might also be "new" in the sense that it was not a value in the dataset. But here we talking about a new or hypothetical single person (i.e., experimental/observational unit).
- $Y_{h(\text{new})}$ = the "new" response (as yet unobserved).
- The best point prediction of $Y_{h(\text{new})}$ is

- Predicts new individual to be the mean for everyone else with X_h

Inference for $Y_{h(\text{new})}$

- Best estimate of prediction error is

$$Y_{h(new\}} - \hat{Y}_h$$

- Note $Y_{h(new\}}$ and $\hat{Y}_h$ are independent
- Variance of the prediction error $\sigma^2\{\text{pred}\}$ is

Inference for $Y_{h(\text{new})}$

- Estimated variance is

$$s^2\{\text{pred}\} = MSE\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right].$$

- Note that

$$\frac{\hat{Y}_h - Y_{h(\text{new})}}{s\{\text{pred}\}} \sim t_{n-2}$$

- The $(1-\alpha)$ prediction interval (PI) for $Y_{h(\text{new})}$ is

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

Prediction of $Y_{h(\text{new})}$

- In the advertising example, again suppose $X_h = 6$.
- The predicted $Y_{h(\text{new})}$ is

$$\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = 0.345 + 1.356 \times 6 = 8.48.$$

- The estimated standard deviation of the prediction error is

$$\begin{aligned} s\{\text{pred}\} &= \sqrt{MSE\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right]} \\ &= \sqrt{0.172} \times \sqrt{1 + \frac{1}{7} + \frac{(6 - 3.486)^2}{22.37}} = 0.495. \end{aligned}$$

- The 95% PI for $Y_{h(\text{new})}$ is

$$\begin{aligned} \hat{Y}_h \pm t(1 - \alpha/2; n-2) s\{\text{pred}\} &= 8.48 \pm 2.571 \times 0.495 \\ &= 8.48 \pm 1.27 = (7.21, 9.75). \end{aligned}$$

Analysis of Variance (ANOVA) Approach

- The idea is to partition the variation into

$$\text{SS Total} = \text{SS Model} + \text{SS Error}$$

- Why partition the variation?
  1. Weigh different sources of variation.
  2. Hypothesis testing.
  3. Comparison of models.

Partitioning Deviation of Each Observation

$$\underbrace{Y_i - \bar{Y}}_{\text{total dev}} = \underbrace{\hat{Y}_i - \bar{Y}}_{\text{dev of fitted from mean}} + \underbrace{Y_i - \hat{Y}_i}_{\text{dev of obs from fitted}} .$$

- If $\{\hat{Y}_i - \bar{Y}\}$ are large in relation to $\{Y_i - \hat{Y}_i\}$, then the regression relation explains (or accounts for) a large proportion of the total variation in $\{Y_i\}$.
- If $\{\hat{Y}_i - \bar{Y}\}$ are small in relation to $\{Y_i - \hat{Y}_i\}$, then the regression relation explains (or accounts for) a small proportion of the total variation in $\{Y_i\}$.

Partitioning Total Sum of Squares

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{\text{SSTO}} = \underbrace{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}_{\text{SSE}}.$$

- SSTO $= \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ is the total sum of squares.
  - A measure of total variation in the data (compare to variance).
- SSR $= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$ is the regression sum of squares.
  - The larger the SSR in relation to SSTO, the larger the proportion of variability in the Y_i's accounted for by the regression relation.
- SSE $= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ is the error sum of squares.
  - The greater the variation of the Y_i's around the fitted regression line, the larger the SSE.

Partitioning Total Sum of Squares

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

where

1. $$\begin{aligned} \text{SSTO} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} Y_i)^2 \\ \text{df} &= n - 1 \end{aligned}$$

2. $$\begin{aligned} \text{SSR} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n}(X_i - \bar{X})^2 = \hat{\beta}_1^2 \left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2\right] \\ \text{df} &= 1 \end{aligned}$$

3. $$\begin{aligned} \text{SSE} &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \text{SSTO} - \text{SSR} \\ \text{df} &= n - 2 \end{aligned}$$

Partitioning Degrees of Freedom

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{\text{df}=n-1} = \underbrace{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}_{\text{df}=1} + \underbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}_{\text{df}=n-2}.$$

- SSTO df $= n - 1$: $\bar{Y}$ is used to estimate μ_Y.
- SSE df $= n - 2$: $\hat{\beta}_0, \hat{\beta}_1$ are used to estimate β_0, β_1.
- Reasons to partition df?
  - Compute MSE and MSR.
  - See ST640.

Expected Mean Squares $E(MSE)$

- Define

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

- Since $SSE/\sigma^2 \sim \chi^2(n-2)$, we have

$$E(MSE) = \sigma^2.$$

Expected Mean Squares $E(MSR)$

- Define $MSR = \frac{SSR}{1}$, recall $SSR = \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2$, we have

$$E(\hat{\beta}_1^2) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \beta_1^2$$

  Why?

- Thus

$$E(MSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^n (X_i - \bar{X})^2.$$

  - Observe that when $\beta_1 = 0$, $E(MSR) = \sigma^2$.

Expected Mean Squares

- Thus for testing $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$, use the test statistic

$$\frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE}$$

- It can be shown that under $H_0 : \beta_1 = 0$,

$$F^* = \frac{MSR}{MSE} \sim F_{1,n-2}.$$

- Thus we can perform an F-test instead of a t-test.
- In fact, $T \sim t_\nu$ then $T^2 \sim F_{1,\nu}$
  - Thus the F-test is equivalent to a two-sided t-test for $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$.

Example SSTO

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \\
&= \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} Y_i)^2 \\
&= 222.03 - \frac{1}{7}(35.50)^2 \\
&= 41.99 \\
\text{df} &= n - 1 = 6
\end{aligned}
$$

Example SSR and SSE

$$
\begin{aligned}
SSR &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\
&= \hat{\beta}_1^2 \left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2\right] \\
&= 1.356^2 \times (107.42 - 24.40^2/7) \\
&= 41.13 \\
\text{df} &= 1 \\
SSE &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\
&= \text{SSTO} - \text{SSR} \\
&= 41.99 - 41.13 = 0.86 \\
\text{df} &= n - 2 = 5
\end{aligned}
$$

General Linear Test Approach

- Consider the full model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad \epsilon_i \sim \text{iid } N(0, \sigma^2)$$

  and obtain SSE(F).

- Consider the reduced model when $\beta_1 = 0$

$$Y_i = \beta_0 + \epsilon_i, \quad \epsilon_i \sim \text{iid } N(0, \sigma^2)$$

  and obtain SSE(R).

- It can be shown that $SSE(F) \leq SSE(R)$ (intuitively, why?)
- In addition, under $H_0 : \beta_1 = 0$,

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} \sim F(df_R - df_F, df_F)$$

Example

- To test $H_0 : \beta_1 = 0$, the F test statistic is

$$F^* = \frac{MSR}{MSE} = \frac{41.13}{0.172} = 239.13$$

- Compare with $F(1, 5)$ and the p-value is

$$P(F(1, 5) > F^*) = P(F(1, 5) > 239.13) < 0.0001.$$

- Same conclusion as in the t test.

ANOVA Table

- Summarize results using an ANOVA table.

Source	SS	df	MS	F
Regression (X)	SSR	1	$SSR/1$	MSR/MSE
Error	SSE	$n-2$	$SSE/n-2$	–
Total	$SSTO$	$n-1$	–	–

- For the advertising example, $n = 7$, then

Source	SS	df	MS	F
Ad expenditure				
Error	0.86			–
Total	41.99		–	–

Coefficient of Determination R^2

- Recall that
  - SSTO measures the variation in Y_i about $\bar{Y}$ (which does not take into account X_i).
  - SSE measures variation in Y_i after accounting for linear relationship between X and Y.
  - SSTO $-$ SSE $=$ SSR is a measure of the reduction of variation due to regression of Y on X.
- Define a coefficient of determination as

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

Coefficient of Determination R^2

- In the advertising example,

$$R^2 = \frac{41.13}{41.99} = 0.9791$$

- Interpret R^2 as the proportion of variation in the Y_i's explained by the linear regression relationship between X and Y
- $0 \leq R^2 \leq 1$.
- Reported as the "multiple R-squared" in R summary output.
- Relation to the sample correlation coefficient for simple linear regression model (only):

$$r = \text{sign}(\hat{\beta}_1)\sqrt{R^2}$$

Can you show that?

Limitations of R^2

- High R^2 does not guarantee that useful predictions can be made.
- Low R^2 does not imply lacks of associations.
  - You can get R^2 near zero even when there is a strong (or perfect) relationship between X and Y.
    - ⋆ Sample correlation only measures **LINEAR** relationship
    - ⋆ E.g., $X \sim N(0,1)$, $Y = \sin(X^2)$, try yourself
  - Outliers.
- Alternative measure: Spearmen correlation (using nonparametric statistics), Lowess R^2 etc.

Correlation Analysis

- Correlation analysis (Section 2.11 of KNNL) is closely related to regression analysis
- Regression analysis
  - One variable is the response Y
  - One variable is the predictor X
  - Model conditional distribution of Y given X
  - Distribution of X is not relevant
  - Y given X and X given Y are not the same
- Correlation analysis
  - Both variables are response variables
  - Want to measure association between two variables
  - $\rho_{X,Y}$ and $\rho_{Y,X}$ are the same

Bivariate Normal Distribution

- The bivariate normal distribution is an example of a joint distribution for two continuous random variables
- We say X and Y have a bivariate normal distribution with parameters $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$ if the probability density is

- Interpretation of parameters
  1. μ_x = mean of X
  2. μ_y = mean of Y
  3. σ_x^2 = variance of X
  4. σ_y^2 = variance of Y
  5. ρ= correlation of X and Y

Definition of the Correlation Coefficient

- $\rho = \text{Cov}(X, Y)/(\sigma_x \sigma_y)$ is called the correlation coefficient
- Recall that $\text{Var}(X) = \mathbb{E}\{(X - \mu_x)^2\}$ and $\text{Cov}(X, Y) =$?

- Properties
  1. $\rho \in [-1, 1]$
  2. $|\rho| = 1$ implies that X and Y have a perfect linear relationship (**perfect correlation**)
  3. Independence of X and Y implies that $\rho = 0$
  4. $\rho = 0$, on the other hand, implies independence (no linear relationship) when (X, Y) is bivariate normal

Bivariate Normal Distribution

- Density is constant on ellipses (contour plots)
- Marginal distributions are normal that

- Conditional distributions are normal that

- Note the relationship of the conditional distribution of Y given $X = x$ to simple linear regression model

Bivariate Normal Distribution

- Two motivations for simple linear regression
  1. Bivariate normal observations
  2. x is fixed (not necessarily normal) and $Y|x$ is normal
- We can relate the bivariate normal and simple linear regression parameters:

Inference for a Correlation Coefficient

- The maximum likelihood estimate is the sample correlation coefficient (Pearson correlation)

  This estimate replaces population quantities with sample quantities

- Test $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$
  - Equivalent to testing $\beta_1 = 0$ in regression
  - $t = r\sqrt{n-1}/\sqrt{1-r^2}$ has t-distribution with $n-2$ df
  - This is exactly the t-test for $H_0 : \beta_1 = 0$

Inference for a Correlation Coefficient

- Remaining inference procedures assume bivariate normality and rely on Fisher's z-transformation

$$Z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

- $Z \sim N(\log((1+\rho)/(1-\rho))/2, 1/(n-3))$
- Var(Z) does not depend on ρ!
- Good approximation when $n > 25$, why?

Inference for a Correlation Coefficient

Construct an $100(1-\alpha)\%$ confidence interval for ρ

- CI for $\log((1+\rho)/(1-\rho))/2$ is

- Obtain an approximate confidence interval for ρ by applying the inverse transformation to the ends of the previous confidence interval

Inference for a Correlation Coefficient

Test $H_0 : \rho = \rho_0$ versus $H_1 : \rho \neq \rho_0$

- The test statistic is

$$\sqrt{n-3}\left(\frac{1}{2}\log\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log\left(\frac{1+\rho_0}{1-\rho_0}\right)\right)$$

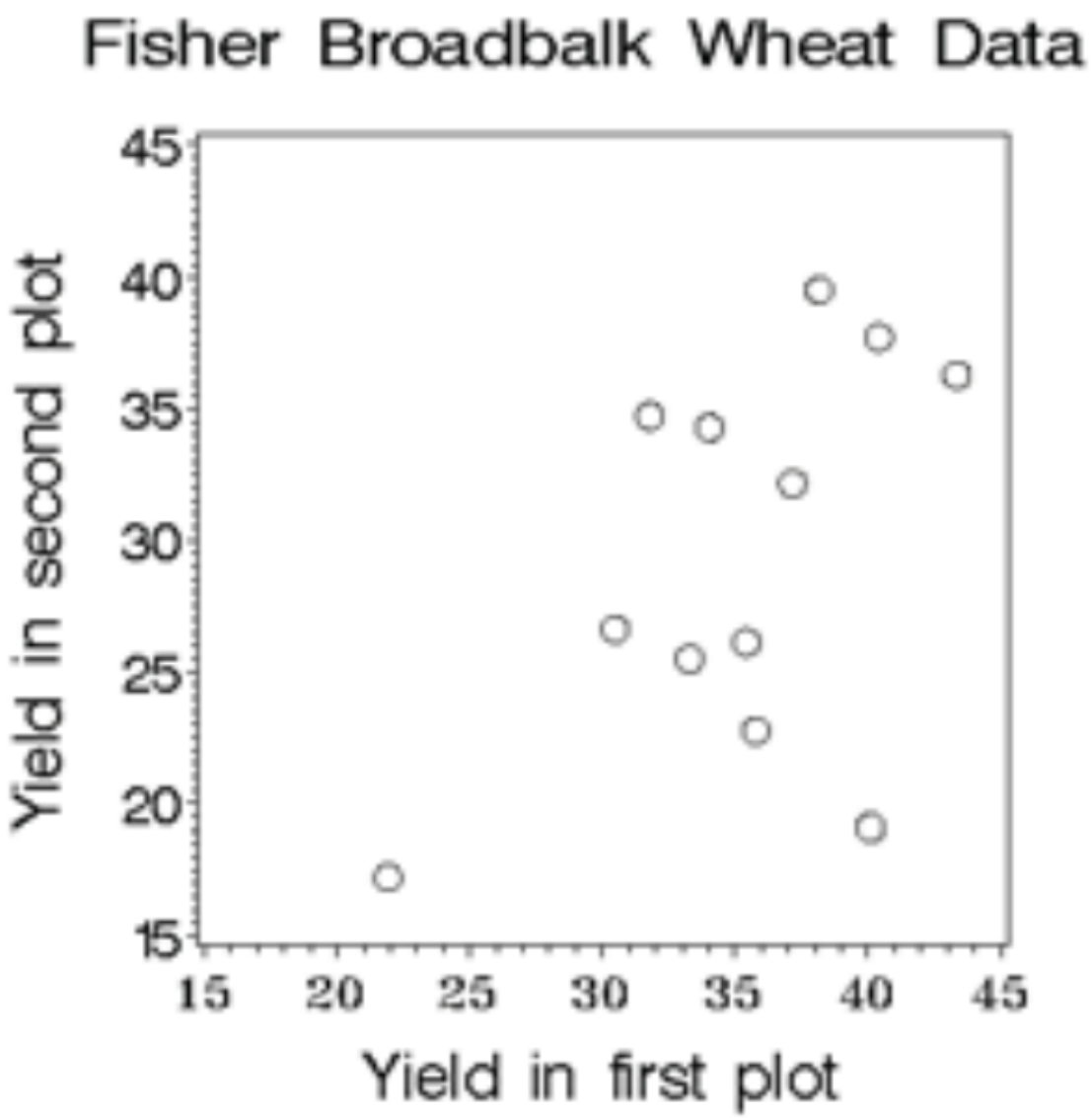
- Obtain the p-value from comparison to a standard normal distribution

Example: Yields of Broadbalk Wheat (bu/acre)

Source: R. A. Fisher, Statistical Methods for Research Workers, 14th ed. page 137.

- Same two plots used in each of n=12 years
  - Plot 1: fertilized with nitrate of soda, X_i=yield in i-th year
  - Plot 2: same amount of N as sulfate of ammonia, Y_i = yield in i-th year

Year	Plot 1	Plot 2
1873	x_1=35.81	y_1=22.75
1874	x_2=38.19	y_2=39.56
1875	x_1=30.50	y_1=26.63
1876	x_2=33.31	y_2=25.50
1877	x_1=40.12	y_1=19.12
1878	x_2=37.19	y_2=32.19
1879	x_1=21.94	y_1=17.25
1880	x_2=34.06	y_2=34.31
1881	x_1=35.44	y_1=26.13
1882	x_2=31.81	y_2=34.75
1883	x_1=43.38	y_1=36.31
1884	x_2=40.44	y_2=37.75


Fisher Broadbalk Wheat Data
Yield in second plot
45
40
35
30
25
20
15
15
20
25
30
35
40
45
Yield in first plot

Grain Example

Summary statistics

- Sample size: $n = 12$
- Sample Means: $\bar{x} = 35.1825$ and $\bar{y} = 29.3541$
- Sums of Squares: $\sum_{i=1}^{12}(x_i - \bar{x})^2 = 346.184$, $\sum_{i=1}^{12}(y_i - \bar{y})^2 = 612.285$
- Sums of Crossproducts: $\sum_{i=1}^{12}(x_i - \bar{x})(y_i - \bar{y}) = 238.5449$
- Sample correlation: $r = 0.518$

- Test $H_0 : \rho = 0$ versus $H_1 : \rho \neq_0$
- $t = (0.518\sqrt{12-2})/(\sqrt{1-0.518^2}) = 2.24$ where $t \sim t_{12-2}$ and p-value is 0.0844
- Conclusion: There is some evidence of a positive correlation in yields, but it is not conclusive. Why?
  - $n = 12$ is a small sample size
  - The pair of yields observed in 1877 is somewhat inconsistent with the pattern observed in other years. Check the accuracy of the 1877 data.

Confidence Interval for ρ

- Apply the Fisher z-transformation

$$z = \frac{1}{2}\log\left(\frac{1+0.518}{1-0.518}\right) = 0.5736$$

$$z_{lower} = 0.5736 - (1.96)\sqrt{1/9} = -0.0797$$

$$z_{upper} = 0.5736 + (1.96)\sqrt{1/9} = 1.2269$$

- Apply the inverse transformation

$$\left(\frac{-1+\exp(2(-0.0797))}{1+\exp(2(-0.0797))}, \frac{-1+\exp(2(1.2269))}{1+\exp(2(1.2269))}\right) \Rightarrow (-0.0795, 0.8417)$$