50.034 - Introduction to Probability and Statistics

Week 13 – Cohort Class

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

CC Week 13 Assigned Readings: (Revise: 11.1, 11.2, 11.3, 11.5)

Outline of Cohort Class

Exercises on the following topics:

- Linear regression and least squares estimators
- Hypothesis testing on regression coefficients

Least Squares Method

Theorem: Suppose we are given n points $(x_1, y_1), \ldots, (x_n, y_n)$. Let $\overline{x} = \frac{x_1+\cdots+x_n}{n}$ and $\overline{y} = \frac{y_1+\cdots+y_n}{n}$. The straight line $y = \beta_0 + \beta_1 x$ that **minimizes** $S = \sum_{i=1}^{n}[y_i - (\beta_1 x_i + \beta_0)]^2$ (i.e. the sum of the squares of the vertical deviations of all points) has coefficients:

$$\beta_1 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2};$$

$$\beta_0 = \overline{y} - \beta_1 \overline{x}.$$

- This line is called the least squares line of the n given points.
- Least squares method: Process of computing β_0, β_1.

Intuition: Treat $S = S(\beta_0, \beta_1)$ as a function in terms of β_0 and β_1.

- We want to minimize $S(\beta_0, \beta_1)$, so compute the partial derivatives $\frac{\partial S}{\partial \beta_0}$, $\frac{\partial S}{\partial \beta_1}$, set them to zero, and solve for β_0, β_1.
- **Theorem:** If $S(\beta_0, \beta_1)$ has a unique critical point, then this critical point must be a global minimum point.

Least Squares Method in Higher Dimensions

- Let $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be n given points in $\mathbb{R}^{k+1}$
- Each point $\mathbf{z}_i$ can be written as the pair $(\mathbf{x}_i, y_i)$ (i.e. split into first k entries and last entry), where $\mathbf{x}_i$ is an element of $\mathbb{R}^k$ and y_i is a real number.

	Points in $\mathbb{R}^2$	**Points in $\mathbb{R}^{k+1}$**
Coordinates:	(x_i, y_i)	$(x_{i,1}, x_{i,2}, \ldots, x_{i,k}, y_i)$
Best-fit:	Least squares line	Least squares hyperplane
Equation:	$y = \beta_0 + \beta_1 x$	$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$
Variables:	y, x	y, $x_1, \ldots, x_k$

Least squares method: Process of computing $\beta_0, \beta_1, \ldots, \beta_k$.

Intuition: $S = S(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \sum_{j=1}^{k} \beta_j x_{i,j})\right]^2$ is treated as a function of $\beta_0, \beta_1, \ldots, \beta_k$ to be minimized.

- Compute the partial derivatives $\frac{\partial S}{\partial \beta_0}, \ldots, \frac{\partial S}{\partial \beta_k}$, set them to zero, and solve for $\beta_0, \ldots, \beta_k$.
- **Theorem:** If $S(\beta_0, \ldots, \beta_k)$ has a unique critical point, then this critical point must be a global minimum point.

Simple Linear Regression

Statistical Model: $Y = \beta_0 + \beta_1 X + E$, where β_0 and β_1 are unknown parameters.

- $\{X_1, \ldots, X_n\}$, $\{Y_1, \ldots, Y_n\}$ are random samples of observable R.V.'s, where the observed pairs $(x_1, y_1), \ldots, (x_n, y_n)$ (for $(X_1, Y_1), \ldots, (X_n, Y_n)$ respectively) are points independently sampled from the joint distribution of (X, Y).

Assumptions:

- E is a normal R.V. with mean 0 and variance σ^2.
- Observed values $X_1 = x_1, \ldots, X_n = x_n$ are known beforehand.
  - i.e. we can treat $x_1, \ldots, x_n$ as constants given to us.
  - Implicit assumption: $x_1, \ldots, x_n$ have no measurement errors.
- $Y_1, \ldots, Y_n$ are conditionally independent given $x_1, \ldots, x_n$.

Simple linear regression is a type of **statistical inference** to determine estimates for the values of β_0 and β_1.

- **Fact:** $\mathbf{E}[Y|X = x] = \beta_0 + \beta_1 x$ (i.e. linear in terms of x).
- $\beta_0 + \beta_1 x$, when treated as a function of x, is the regression function of Y on X, or more simply, the regression of Y on X. The coefficients β_0, β_1 are called regression coefficients.

Least square estimators

$Y = \beta_0 + \beta_1 X + E$ is our statistical model, where $E \sim N(0, \sigma^2)$. Let $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$ be the M.L.E.'s of $\beta_0, \beta_1, \sigma^2$ respectively.

- **Note:** The conditional distribution of Y given $X = x$ is the normal distribution with mean $\beta_0 + \beta_1 x$ and variance σ^2.

Theorem: Given $X_1 = x_1, \ldots, X_n = x_n$, the M.L.E.'s of $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$ are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \overline{Y}_n)(x_i - \overline{x}_n)}{\sum_{i=1}^n (x_i - \overline{x}_n)^2};$$

$$\hat{\beta}_0 = \overline{Y}_n - \hat{\beta}_1 \overline{x}_n;$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2;$$

where $\overline{Y}_n$ is the sample mean of $\{Y_1, \ldots, Y_n\}$, and $\overline{x}_n = \frac{x_1 + \cdots + x_n}{n}$.

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are called the least square estimators of β_0 and β_1.
  - Same expressions as β_0 and β_1 in the **least squares method**.
- $\hat{\sigma}^2$ equals $\frac{1}{n}$ times the observed sum of the squares of the vertical deviations (except that β_0, β_1 are replaced by $\hat{\beta}_0, \hat{\beta}_1$).

Joint distribution of least square estimators

$\hat\beta_0$ and $\hat\beta_1$ are estimators, so in particular, they are R.V.'s.

- It makes sense to consider the joint distribution of $\hat\beta_0$ and $\hat\beta_1$.

Recall: By assumption, the observed values $x_1, \ldots, x_n$ are known beforehand. Let $\overline{x}_n = \frac{x_1+\cdots+x_n}{n}$, and define the real number

$$s_{\mathbf{x}} = \sqrt{\sum_{i=1}^{n} (x_i - \overline{x}_n)^2}.$$

Theorem: The joint distribution of $\hat\beta_0$ and $\hat\beta_1$ is a **bivariate normal distribution**.

- $\hat\beta_0$ has mean β_0 and variance $\sigma^2\left(\frac{1}{n} + \frac{\overline{x}_n^2}{s_{\mathbf{x}}^2}\right)$.
- $\hat\beta_1$ has mean β_1 and variance $\frac{\sigma^2}{s_{\mathbf{x}}^2}$.
- The covariance of $\hat\beta_0$ and $\hat\beta_1$ is $\mathrm{cov}(\hat\beta_0, \hat\beta_1) = -\frac{\overline{x}_n \sigma^2}{s_{\mathbf{x}}^2}$.
  - Thus, the correlation is $\frac{\mathrm{cov}(\hat\beta_0,\hat\beta_1)}{\sqrt{\mathrm{var}(\hat\beta_0)\mathrm{var}(\hat\beta_1)}} = \frac{-\overline{x}_n}{\sqrt{\frac{1}{n}+\frac{\overline{x}_n^2}{s_{\mathbf{x}}^2}}}$.

Corollary: $\hat\beta_0$, $\hat\beta_1$ are **unbiased** estimators of β_0, β_1 respectively.

- "$\hat\beta_i$ is unbiased" means "$\mathbf{E}[\hat\beta_i] = \beta_i$".

Exercise 1 (15 mins)

Suppose we carry out a simple linear regression on the statistical model $Y = \beta_0 + \beta_1 X + E$, where $(x_1, y_1), \ldots, (x_8, y_8)$ are sampled from the joint distribution of (X, Y), whose coordinates are:

i	x_i	y_i
1	1.0	6.9
2	2.0	10.8
3	3.0	9.3
4	4.0	7.8
5	5.0	-0.7
6	6.0	-9.2
7	7.0	-22.1
8	8.0	-37.7

1. What are the least squares estimators of β_0 and β_1?
2. What are the least squares estimates of β_0 and β_1?
3. Assume $\mathrm{var}(E) = \sigma^2$. What is M.L.E. (estimate) of σ^2?

Exercise 1 - Solution

1. By assumption, the values of $x_1, \ldots, x_8$ are known beforehand and can be treated as constants. Note that $\overline{x}_8 = \frac{x_1 + \cdots + x_8}{8} = 4.5$.

- $\sum_{i=1}^{8} (x_i - \overline{x}_8)^2 = 2\left(3.5^2 + 2.5^2 + 1.5^2 + 0.5^2\right) = 42.$
- $x_i - \overline{x}_8 = i - 4.5 \quad$ (for each i).

Thus the least squares estimators of β_1 and β_0 are:

$$\hat{\beta}_1(Y_1, \ldots, Y_8) = \frac{1}{42} \sum_{i=1}^{8} (Y_i - \overline{Y}_8)(i - 4.5);$$

$$\begin{aligned} \hat{\beta}_0(Y_1, \ldots, Y_8) &= \overline{Y}_8 - 4.5 \hat{\beta}_1(Y_1, \ldots, Y_8) \\ &= \overline{Y}_8 - \frac{3}{28} \sum_{i=1}^{8} (Y_i - \overline{Y}_8)(i - 4.5); \end{aligned}$$

where $\overline{Y}_8 = \frac{Y_1 + \cdots + Y_8}{8}$.

Exercise 1 - Solution

2. Let $\mathbf{y} = (6.9, 10.8, 9.3, 7.8, -0.7, -9.2, -22.1, -37.7)$. Note also that $\frac{y_1 + \cdots + y_8}{8} = -4.3625$. Given $\mathbf{y}$, the least square estimates are

$$\hat{\beta}_1(\mathbf{y}) = \frac{1}{42} \sum_{i=1}^{8} (Y_i + 4.3625)(i - 4.5) = -6.4369;$$

$$\hat{\beta}_0(\mathbf{y}) = -4.3625 - 4.5(-6.4369) = 24.6036;$$

3. Using $\hat{\beta}_1(\mathbf{y}) = -6.4369$ and $\hat{\beta}_0(\mathbf{y}) = 24.6036$ from the previous part, we compute that the maximum likelihood estimate of σ^2 is

$$\begin{aligned}
\hat{\sigma}^2(\mathbf{y}) &= \tfrac{1}{8} \sum_{k=1}^{8} (y_k - \hat{\beta}_0(\mathbf{y}) - \hat{\beta}_1(\mathbf{y}) x_k)^2 \\
&= \tfrac{1}{8} \sum_{k=1}^{8} (y_k - 24.6036 + 6.4369k)^2 \\
&\approx 51.7427.
\end{aligned}$$

Using simple linear regression for prediction

Given the observed values $Y_1 = y_1, \dots, Y_n = y_1$, we obtain a "best-fit line" model $Y = \hat{\beta}_0 + \hat{\beta}_1 X$, where $\hat{\beta}_0 = \hat{\beta}_0(y_1, \dots, y_n)$ and $\hat{\beta}_1 = \hat{\beta}_1(y_1, \dots, y_n)$ are real numbers (estimates of β_0, β_1).

- This model is called a simple linear regression model.

Prediction: Given **any** "new" value $X = x$, our simple linear regression model gives the prediction $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

- $\mathbf{E}[(Y - \hat{y})^2]$ is called the mean squared error (M.S.E.) of $\hat{y}$.

Theorem: In simple linear regression, the prediction $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ (given $X = x$) has the **mean squared error**

$$\mathbf{E}[(Y - \hat{y})^2] = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x - \overline{x}_n)}{s_{\mathbf{x}}^2}\right],$$

where $\overline{x}_n = \frac{x_1 + \cdots + x_n}{n}$ is the average of the given observed values $x_1, \dots, x_n$ of the predictor variables, and $s_{\mathbf{x}} = \sqrt{\sum_{i=1}^n (x_i - \overline{x}_n)^2}$.

Interpretation:

- The M.S.E. becomes smaller and closer to σ^2 with a larger n.
- The M.S.E. is larger as x becomes further away from $\overline{x}_n$.

Distribution of M.L.E. of σ^2

Recall: In a simple linear regression model $Y = \hat{\beta}_0 + \hat{\beta}_1 X$, we know $(\hat{\beta}_0, \hat{\beta}_1)$ has a bivariate normal distribution, and Y conditioned on $X = x$ is normal with mean $\beta_0 + \beta_1 x$ and variance σ^2.

- **Recall:** $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$.

Important Theorem: If $n \geq 3$, then $\hat{\sigma}^2$ is independent of $(\hat{\beta}_0, \hat{\beta}_1)$, and $\frac{n\hat{\sigma}^2}{\sigma^2}$ has the χ^2 distribution with $n - 2$ degrees of freedom.

Interpretation:

- To estimate the value of the unknown parameter σ^2, we can use the estimator $\hat{\sigma}^2$ to compute an estimate $\hat{\sigma}^2(y_1, \ldots, y_n)$.
- Even though $\hat{\sigma}^2$ is computed in terms of the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, this theorem says that any estimate of $\hat{\sigma}^2$ is actually independent of the corresponding estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$, i.e. it does not matter what the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ are.

Exercise 2 (20 mins)

Suppose we carry out a simple linear regression on the statistical model $Y = \beta_0 + \beta_1 X + E$, where $E \sim N(0, \sigma^2)$, and where $(x_1, y_1), \ldots, (x_4, y_4)$ are sampled from the joint distribution of (X, Y), whose coordinates are given as follows:

i	x_i	y_i
1	1.0	1.1
2	3.0	2.9
3	6.0	6.1
4	10	10.1

1. What is the corresponding simple linear regression model?
2. What is the maximum likelihood estimate of σ^2?
3. Using your simple linear regression model, predict the value of Y given $X = 7$.
4. What is the mean squared error of your prediction?
5. Let $\hat{\sigma}^2$ be the maximum likelihood estimator of σ^2. What is the probability $\Pr(\hat{\sigma}^2 \leq 0.6\sigma^2)$?

Useful values for the χ^2 distribution

					p				
.50	.60	.70	.75	.80	.90	.95	.975	.99	.995
.4549	.7083	1.074	1.323	1.642	2.706	3.841	5.024	6.635	7.879
1.386	1.833	2.408	2.773	3.219	4.605	5.991	7.378	9.210	10.60
2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	11.34	12.84
3.357	4.045	4.878	5.385	5.989	7.779	9.488	11.14	13.28	14.86
4.351	5.132	6.064	6.626	7.289	9.236	11.07	12.83	15.09	16.75
5.348	6.211	7.231	7.841	8.558	10.64	12.59	14.45	16.81	18.55
6.346	7.283	8.383	9.037	9.803	12.02	14.07	16.01	18.48	20.28
7.344	8.351	9.524	10.22	11.03	13.36	15.51	17.53	20.09	21.95
8.343	9.414	10.66	11.39	12.24	14.68	16.92	19.02	21.67	23.59
9.342	10.47	11.78	12.55	13.44	15.99	18.31	20.48	23.21	25.19
10.34	11.53	12.90	13.70	14.63	17.27	19.68	21.92	24.72	26.76
11.34	12.58	14.01	14.85	15.81	18.55	21.03	23.34	26.22	28.30
12.34	13.64	15.12	15.98	16.98	19.81	22.36	24.74	27.69	29.82
13.34	14.69	16.22	17.12	18.15	21.06	23.68	26.12	29.14	31.32
14.34	15.73	17.32	18.25	19.31	22.31	25.00	27.49	30.58	32.80
15.34	16.78	18.42	19.37	20.47	23.54	26.30	28.85	32.00	34.27
16.34	17.82	19.51	20.49	21.61	24.77	27.59	30.19	33.41	35.72
17.34	18.87	20.60	21.60	22.76	25.99	28.87	31.53	34.81	37.16
18.34	19.91	21.69	22.72	23.90	27.20	30.14	32.85	36.19	38.58
19.34	20.95	22.77	23.83	25.04	28.41	31.41	34.17	37.57	40.00
20.34	21.99	23.86	24.93	26.17	29.62	32.67	35.48	38.93	41.40
21.34	23.03	24.94	26.04	27.30	30.81	33.92	36.78	40.29	42.80
22.34	24.07	26.02	27.14	28.43	32.01	35.17	38.08	41.64	44.18
23.34	25.11	27.10	28.24	29.55	33.20	36.42	39.36	42.98	45.56
24.34	26.14	28.17	29.34	30.68	34.38	37.65	40.65	44.31	46.93
29.34	31.32	33.53	34.80	36.25	40.26	43.77	46.98	50.89	53.67

Exercise 2 - Solution

1. By assumption, the values of $x_1, \ldots, x_8$ are known beforehand and can be treated as constants. Note that $\overline{x}_4 = \frac{x_1 + \cdots + x_4}{4} = 5$. Note also that $\overline{y}_4 = \frac{y_1 + \cdots + y_4}{4} = 5.05$.

- $\sum_{i=1}^{4} (x_i - \overline{x}_4)^2 = (-4)^2 + (-2)^2 + 1^2 + 5^2 = 46$.

Thus the least squares estimators of β_1 and β_0 are:

$$\begin{aligned}
\hat{\beta}_1(y_1, \ldots, y_4) &= \frac{1}{46}((-3.95)(-4) + (-2.15)(-2) + (1.05) + (5.05)(5)) \\
&\approx 1.0087 \\
\hat{\beta}_0(y_1, \ldots, y_4) &= 5.05 - 5\hat{\beta}_1(y_1, \ldots, y_4) \\
&\approx 5.05 - 5(1.0087) = 0.0065.
\end{aligned}$$

Therefore, the corresponding simple linear regression model is $Y = 0.0065 + 1.0087X$.

Exercise 2 - Solution

2. Let $\mathbf{y} = (1.1, 2.9, 6.1, 10.1)$. Using $Y = 0.0065 + 1.0087X$ from the previous part, we compute that the M.L.E. (estimate) of σ^2 is

$$\begin{aligned}\hat{\sigma}^2(\mathbf{y}) &= \tfrac{1}{4}\sum_{k=1}^{4}(y_k - \hat{\beta}_0(\mathbf{y}) - \hat{\beta}_1(\mathbf{y})x_k)^2 \\ &= \tfrac{1}{4}\sum_{k=1}^{4}(y_k - 0.0065 - 1.0087k)^2 \\ &\approx 0.006630.\end{aligned}$$

3. Our prediction, given $X = 7$, is $\hat{y} = 0.0065 + 1.0087(7) = 7.0674$.

4. From part 1., $\overline{x}_4 = 5$ and $\sum_{i=1}^{4}(x_i - \overline{x}_4)^2 = 46$, thus the mean squared error of $\hat{y}$ is $\sigma^2\left[1 + \frac{1}{4} + \frac{7-5}{46}\right] = \frac{119}{92}\sigma^2 \approx 1.2935\sigma^2$.

Exercise 2 - Solution

5. $Z = \frac{4\hat{\sigma}^2}{\sigma^2}$ has the χ^2 distribution with 2 degrees of freedom. Note also that

$$\Pr(\hat{\sigma}^2 \leq 0.6\sigma^2) = \Pr(\frac{4\hat{\sigma}^2}{\sigma^2} \leq 2.4) = \Pr(Z \leq 2.4)$$

Using the table of values for the χ^2 distribution, we note that $\Pr(Z \leq 2.408) = 0.7$. Therefore, $\Pr(\hat{\sigma}^2 \leq 0.6\sigma^2) \approx 0.7$.

					p				
.50	.60	.70	.75	.80	.90	.95	.975	.99	.995
.4549	.7083	1.074	1.323	1.642	2.706	3.841	5.024	6.635	7.879
1.386	1.833	$\boxed{2.408}$	2.773	3.219	4.605	5.991	7.378	9.210	10.60
2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	11.34	12.84
3.357	4.045	4.878	5.385	5.989	7.779	9.488	11.14	13.28	14.86
4.351	5.132	6.064	6.626	7.289	9.236	11.07	12.83	15.09	16.75
5.348	6.211	7.231	7.841	8.558	10.64	12.59	14.45	16.81	18.55
6.346	7.283	8.383	9.037	9.803	12.02	14.07	16.01	18.48	20.28
7.344	8.351	9.524	10.22	11.03	13.36	15.51	17.53	20.09	21.95
8.343	9.414	10.66	11.39	12.24	14.68	16.92	19.02	21.67	23.59
9.342	10.47	11.78	12.55	13.44	15.99	18.31	20.48	23.21	25.19
10.34	11.53	12.90	13.70	14.63	17.27	19.68	21.92	24.72	26.76
11.34	12.58	14.01	14.85	15.81	18.55	21.03	23.34	26.22	28.30
12.34	13.64	15.12	15.98	16.98	19.81	22.36	24.74	27.69	29.82
13.34	14.69	16.22	17.12	18.15	21.06	23.68	26.12	29.14	31.32
14.34	15.73	17.32	18.25	19.31	22.31	25.00	27.49	30.58	32.80
15.34	16.78	18.42	19.37	20.47	23.54	26.30	28.85	32.00	34.27

Regression coefficients and t-distributions

Recall: β_0 and β_1 are called regression coefficients.

- Let $\overline{x}_n = \frac{x_1 + \cdots + x_n}{n}$, and let $\overline{Y}_n = \frac{Y_1 + \cdots + Y_n}{n}$.
- Let the M.L.E.'s of β_0 and β_1 be $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively.
- Let $s_{\mathbf{x}} = \sqrt{\sum_{i=1}^{n} (x_i - \overline{x}_n)^2}$; $\sigma' = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$.
  - **Recall:** $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$; hence $(\sigma')^2 = \frac{n}{n-2} \hat{\sigma}^2$.
- **Note:** The joint distribution of $(\hat{\beta}_0, \hat{\beta}_1)$ is bivariate normal.
  - So any (non-zero) linear combination of $\hat{\beta}_0$ and $\hat{\beta}_1$ is normal.

Theorem: Given any (non-zero) linear combination $c_0\beta_0 + c_1\beta_1$, its M.L.E. is $c_0\hat{\beta}_0 + c_1\hat{\beta}_1$, which has a normal distribution with mean $c_0\beta_0 + c_1\beta_1$ and variance $\sigma^2\left(\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_{\mathbf{x}}^2}\right)$.

Important Theorem: Let c_0, c_1 be constants, not both zero, and suppose that $\mathbf{E}[c_0\beta_0 + c_1\beta_1] = c$. If $n \geq 3$, then the statistic

$$T = \left[\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_{\mathbf{x}}^2}\right]^{-0.5} \left(\frac{c_0\hat{\beta}_0 + c_1\hat{\beta}_1 - c}{\sigma'}\right)$$

has the t-distribution with $(n-2)$ degrees of freedom.

t-tests for linear combinations of regression coefficients

Same as before: $Y_1, \ldots, Y_n$ are observable R.V.'s; $x_1, \ldots, x_n$ are given values. Let $\overline{x}_n = \frac{x_1 + \cdots + x_n}{n}$.

- Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the M.L.E.'s of β_0 and β_1 respectively.
- Let $s_{\mathbf{x}} = \sqrt{\sum_{i=1}^n (x_i - \overline{x}_n)^2}$; $\sigma' = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$.

Important consequences of theorem on the previous slide: Assume $n \geq 3$. Let c_0, c_1, c be constants such that c_0, c_1 are not both zero, and such that $\mathbf{E}[c_0\beta_0 + c_1\beta_1] = c$. Define the statistic

$$T = \left[\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_{\mathbf{x}}^2}\right]^{-0.5} \left(\frac{c_0\hat{\beta}_0 + c_1\hat{\beta}_1 - c}{\sigma'}\right).$$

- If $\mathcal{H}$ is a hypothesis test with $H_0 : c_0\beta_0 + c_1\beta_1 \leq c$, test statistic T, and rejection region $[k, \infty)$, then $\mathcal{H}$ is a t-test.
- If $\mathcal{H}$ is a hypothesis test with $H_0 : c_0\beta_0 + c_1\beta_1 \geq c$, test statistic T, and rejection region $(-\infty, k]$, then $\mathcal{H}$ is a t-test.
- If $\mathcal{H}$ is a hypothesis test with $H_0 : c_0\beta_0 + c_1\beta_1 = c$, test statistic $|T|$, and rejection region $[k, \infty)$, then $\mathcal{H}$ is a t-test.

Exercise 3 (30 mins)

(**Same four sample points as in Example 2**)
Suppose we carry out a simple linear regression on the statistical model $Y = \beta_0 + \beta_1 X + E$, where $(x_1, y_1), \ldots, (x_4, y_4)$ are sampled from the joint distribution of (X, Y), whose coordinates are given as follows:

i	x_i	y_i
1	1.0	1.1
2	3.0	2.9
3	6.0	6.1
4	10	10.1

Define the statistic $T = \frac{\hat{\beta}_0 - 0.448}{\sqrt{\sum_{i=1}^{4}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}}$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the maximum likelihood estimators of β_0 and β_1 respectively.
Let $\mathcal{H} = \{\mathcal{H}_c\}_{c \in \mathbb{R}}$ be a collection of hypothesis tests, where each $\mathcal{H}_c$ has null hypothesis $H_0 : \beta_0 = 0.448$, alternative hypothesis $H_1 : \beta_1 \neq 0.448$, test statistic $|T|$, and rejection region $[c, \infty)$.
Determine the p-value of $\mathcal{H}$.

Useful values for the t-distribution

Table of the t Distribution

If X has a t distribution with m degrees of freedom, the table gives the value of x such that $\Pr(X \le x) = p$.

m	$p = .55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845

Exercise 3 - Solution

Let $\mathbf{y} = (1.1, 2.9, 6.1, 10.1)$. From Exercise 2, we already have:

$$\overline{x}_4 = \tfrac{x_1+\cdots+x_4}{4} = 5; \qquad \textstyle\sum_{i=1}^4 (x_i - \overline{x}_4)^2 = 46;$$

$$\hat{\beta}_0(\mathbf{y}) \approx 0.0065; \qquad \hat{\beta}_1(\mathbf{y}) \approx 1.0087;$$

$$\hat{\sigma}^2(\mathbf{y}) = \tfrac{1}{4}\sum_{i=1}^{4}(y_i - \hat{\beta}_0(\mathbf{y}) - \hat{\beta}_1(\mathbf{y})x_i)^2 \approx 0.006630.$$

Recall: Let c_0, c_1 be constants, not both zero, and suppose that $\mathbf{E}[c_0\beta_0 + c_1\beta_1] = c$. If $n \geq 3$, then the statistic

$$T = \left[\tfrac{c_0^2}{n} + \tfrac{(c_0\overline{x}_n - c_1)^2}{s_{\mathbf{x}}^2}\right]^{-0.5}\left(\frac{c_0\hat{\beta}_0 + c_1\hat{\beta}_1 - c}{\sigma'}\right)$$

has the t-distribution with $(n-2)$ degrees of freedom, where $s_{\mathbf{x}} = \sqrt{\sum_{i=1}^n (x_i - \overline{x}_n)^2}$ and $\sigma' = \sqrt{\frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$.

Exercise 3 - Solution (continued)

Substitute $c_0 = 1, c_1 = 1, c = 0.448$, and our given observed values, so that $\mathbf{E}[\beta_0] = 0.5$, $s_{\mathbf{x}} = \sqrt{46}$, we get that

$$\frac{\hat{\beta}_0 - 0.448}{\sqrt{\frac{1}{2}\sum_{i=1}^{4}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}\sqrt{\frac{1}{4} + \frac{5^2}{46}}} = \sqrt{\frac{184}{73}}T.$$

has the t-distribution with 2 degrees of freedom.

Given $\mathbf{y}$, we check that the observed value of T is

$$T(\mathbf{y}) = \frac{\hat{\beta}_0(\mathbf{y}) - 0.448}{\sqrt{4\hat{\sigma}^2(\mathbf{y})}} = \frac{0.0065 - 0.448}{\sqrt{4(0.006630)}} \approx -2.7111$$

Thus $|\sqrt{\frac{184}{73}}T(\mathbf{y})| = 4.3042$.

Exercise 3 - Solution (continued)

So far: $\left|\sqrt{\frac{184}{73}}\,T(\mathbf{y})\right| = 4.3042.$

Using a table of values for the t-distribution, we check that 4.3042 is approximately the 97.5-th percentile of the t-distribution with 2 degrees of freedom.

Since $2(1-0.975) = 0.05$, we conclude that the p-value of $\mathcal{H}$ is 0.05.

Table of the *t* Distribution

If X has a t distribution with m degrees of freedom, the table gives the value of x such that $\Pr(X \le x) = p$.

m	$p = .55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	$\boxed{4.303}$	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106

Summary

Exercises on the following topics:

- Linear regression and least squares estimators
- Hypothesis testing on regression coefficients

Reminder:
The **Final Exam** will be held on 3rd May (Friday), 9–11am, at the **Indoor Sports Hall 2** (61.106).

- Tested on all materials covered in this course
  - Lectures 1–24 and Cohort classes weeks 1–13.
- 1 piece of A4-sized double-sided **handwritten** cheat sheet is allowed for the final exam.
  - A formula sheet similar to that given in the mid-term exam will also be provided. Details of this formula sheet will be announced soon.