STAT2401: Analysis of Experiments
Introduction & Review

Dr Darfiana Nur

Dept. of Math. & Stat.

February 27, 2024

Aims of this week

- Unit administration (using the LMS)
- Aim 1 Motivation Examples
- Aim 2 Background knowledge
- Aim 2.1 Quick Review 1: Random Variables, Normal
- Aim 2.2 Quick Review 2: Populations and Samples
- Aim 2.3 Quick Review 3: Distributions, Expectations and Variance (Normal distribution, Law of Large Numbers (LLN), Central Limit Theorem (CLT) and t distributions)
- Aim 2.4 Quick Review 4: Simple matrix algebra in R (independent reading, independent learning) [eg create an Rmd file, explore the R commands within the slides)

References: Chapters 1-2, An Introduction to Statistical Learning with

Topics covered in this unit

- Simple Linear Regression (SLR) Model
- Multiple Linear Regression (MLR) Model
- Analysis of Variance (ANOVA) Model
- Analysis of Covariance (ANCOVA) Model
- Model Diagnostics
- Model Selection

References: Chapters 3 and 6, An Introduction to Statistical Learning with Applications in R (James ET AL, 2nd edition, 2023, https://www.statlearning.com/) ; A Modern Approach to Regression with R (Sheather, 2009)

What you need to know

Analyse data with R.

Please read the unit outline thoroughly.

The questions in the exam, assignments and tests are related to the implementation, including interpretation of the R-output

- two online tests (10% each) in Weeks 6 and 10; open Friday at 8am, due by the same day at 8pm; open-book
- two assignments (15% each), due in Week 7 (available in Week 4) and due in Week 11 (available in Week 7)
- final examination (50%): LMS examination that will be held during the examination period.
    1. Depend on the circumstances, the examination might be conducted in a computer laboratory/BYOD setting as a face-to-face school-based exam, subject to available facilities (UniApps) or as MS Teams online on LMS.
    2. Students will be informed about the detail of the examination in due course.
- we introduce theorems to explain the implementation and R-output
- proofs are not assessed in the exam or tests.

Statistical Learning

- *Statistical learning* refers to a vast set of tools for *understanding data*.
- These tools can be classified as *supervised* or *unsupervised*.
- *Supervised Statistical Learning* involves building a statistical model for predicting, or estimating, an output based on one or more inputs. Problems of this nature occur in fields as diverse as business, medicine, astrophysics, and public policy.
- *Linear regression* is an approach for *Supervised Statistical Learning*.
- With *Unsupervised Statistical Learning*, there are inputs but no supervising output; nevertheless we can learn relationships and structure from such data.
- To provide an illustration of some applications of statistical learning, we briefly discuss two real-world data sets.

Aim 1: Motivation Examples: Example 1 Wage Data

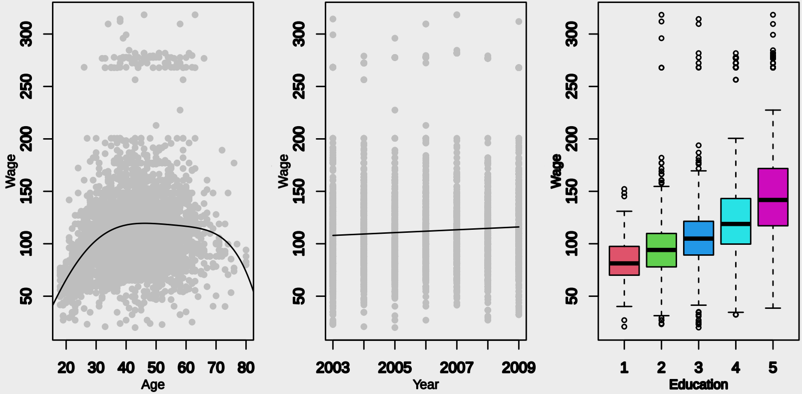
- In this application (which we refer to as the Wage data set), we examine a number of factors that relate to wages for a group of males from the Atlantic region of the United States.

```
> load("Wage.RData")
> str(Wage)
'data.frame': 3000 obs. of  11 variables:
 $ year      : int  2006 2004 2003 2003 2005 2008 2009 2008 2006 2004 ...
 $ age       : int  18 24 45 43 50 54 44 30 41 52 ...
 $ maritl    : Factor w/ 5 levels "1. Never Married",..: 1 1 2 2 4 2 2 1 1 2 .
 $ race      : Factor w/ 4 levels "1. White","2. Black",..: 1 1 1 3 1 1 4 3 2
 $ education : Factor w/ 5 levels "1. < HS Grad",..: 1 4 3 4 2 4 3 3 3 2 ...
 $ region    : Factor w/ 9 levels "1. New England",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ jobclass  : Factor w/ 2 levels "1. Industrial",..: 1 2 1 2 2 2 1 2 2 2 ...
 $ health    : Factor w/ 2 levels "1. <=Good","2. >=Very Good": 1 2 1 2 1 2 2
 $ health_ins: Factor w/ 2 levels "1. Yes","2. No": 2 2 1 1 1 1 1 1 1 1 ...
 $ logwage   : num  4.32 4.26 4.88 5.04 4.32 ...
 $ wage      : num  75 70.5 131 154.7 75 ...
```

Example 1: Wage Data

- We wish to understand the association between an employee's age and education, as well as the calendar year, on his wage.

```
> par(mfrow=c(1,3))
> plot(wage~age,data=Wage)
> plot(wage~year,data=Wage)
> plot(wage~education,data=Wage)
```



Example 1: Wage Data

- The left-hand panel, the display of wage versus age
  - There is evidence that wage increases with age but then decreases again after approximately age 60. The line, which provides an estimate of the average wage for a given age, makes this trend clearer.
- The center and right-hand panels display wage as a function of both year and education,
  - year and education are associated with wage
  - Wages increase by approximately $10,000, in a roughly linear (or straight-line) fashion, between 2003 and 2009, though this rise is very slight relative to the variability in the data.
  - Wages are also typically greater for individuals with higher education levels: men with the lowest education level (1) tend to have substantially lower wages than those with the highest education level (5).
- The most accurate prediction of a given man's wage will be obtained by combining his age, his education, and the year.

Example 2: Stock Market Data (Smarket)

- We examine a stock market data set that contains the daily movements in the Standard & Poor's 500 (S&P) stock index over a 5-year period between 2001 and 2005. We refer to this as the Smarket data.

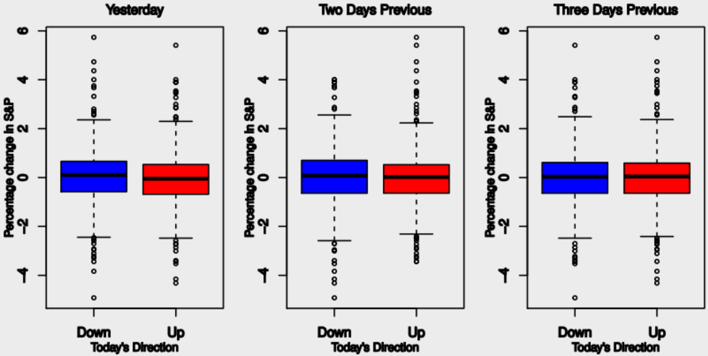
```
> Smarket = read.csv("Smarket.csv",header=TRUE)
> str(Smarket)
```

```
'data.frame': 1250 obs. of  9 variables:
 $ Year     : int  2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 ...
 $ Lag1     : num  0.381 0.959 1.032 -0.623 0.614 ...
 $ Lag2     : num  -0.192 0.381 0.959 1.032 -0.623 ...
 $ Lag3     : num  -2.624 -0.192 0.381 0.959 1.032 ...
 $ Lag4     : num  -1.055 -2.624 -0.192 0.381 0.959 ...
 $ Lag5     : num  5.01 -1.055 -2.624 -0.192 0.381 ...
 $ Volume   : num  1.19 1.3 1.41 1.28 1.21 ...
 $ Today    : num  0.959 1.032 -0.623 0.614 0.213 ...
 $ Direction: chr  "Up" "Up" "Down" "Up" ...
```

Example 2: Stock Market Data (`Smarket`)

- The goal is to predict whether the index will increase or decrease on a given day using the past 5 days' percentage changes in the index.

```
> par(mfrow=c(1,3))
> plot(Lag1~Direction,data=Smarket)
> plot(Lag2~Direction,data=Smarket)
> plot(Lag3~Direction,data=Smarket)
```



Example 2: Stock Market Data (`Smarket`)

- The left-hand panel displays two boxplots of the previous day's percentage changes in the stock index.
  - One for the 648 days for which the market increased on the subsequent day, and one for the 602 days for which the market decreased.
  - The two plots look almost identical, suggesting that there is no simple strategy for using yesterday's movement in the S&P to predict today's returns.
- The remaining panels, which display boxplots for the percentage changes 2 and 3 days previous to today
  - These plots indicate little association between past and present returns. Of course, this lack of pattern is to be expected:
  - in the presence of strong correlations between successive days' returns, one could adopt a simple trading strategy to generate profits from the market.

Example 2: Stock Market Data (Smarket)

- Interestingly, there are hints of some weak trends in the data that suggest that, at least for this 5-year period, it is possible to correctly predict the direction of movement in the market approximately 60% of the time.

	Actual	
Predict	Down	Up
Down	35	35
Up	76	106

```
> (36+106)/(36+106+76+35)
[1] 0.5612648
> 106/(106+76)
[1] 0.5824176
```

Aim 2: Background Knowledge

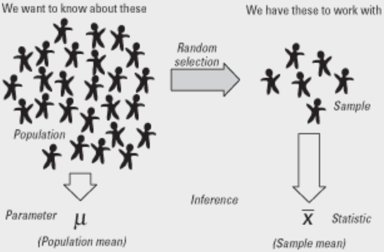
Most people doing this unit should be aware of the following:

- Random variables (In particular, Gaussian/Normal Random Variable) and their distributions
- Populations and samples
- Statistical inference based on the theorem of random variables including;
    - Distributions, expectations and variances
    - Law of large numbers
    - Central limit theorem
    - Tests of significance and confidence intervals

Aim 2.1 Random Variables

- A random variable is a quantity Y that depends on the outcome of an experiment.
- Traditionally written with a capital letter, e.g. Y
- Two major types of random variables are "discrete" and "continuous" Random variables.
- A discrete random variable has a probability mass function, pmf.
- A continuous random variable has a probability density function, pdf.
- Both discrete and continuous random variables have cumulative distribution functions, cdf.

Aim 2.2 Populations and Samples



From www.cliffsnotes.com

Our interests are not limited to

- Population mean μ
- Population variance σ^2
- Population median

Aim 2.3: Distributions, Expectations and Variances

- Let Y be a random variable.
- The cumulative distribution function of Y is $F(y) = P(Y \leq y)$
- When Y is continuous, and it takes values from $(-\infty, \infty)$, then
  - It has a pdf, $f(y) = \frac{d}{dy}F(y)$
  - The expecated value (or expectation) is $\mu = E(Y) = \int_{-\infty}^{\infty} yf(y)dy$
  - The variance is $\sigma^2 = Var(Y) = E((Y-\mu)^2) = E[Y^2] - (E[Y])^2$ where $E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y)dy$
- When Y is discrete, and it take values from the set $\{y_1, \ldots, y_m\}$
  - It has a pmf, $P(Y = y_j)$, $j = 1, \ldots, m$
  - The expecated value (or expectation) is $\mu = E(Y) = \sum_{j=1}^{m} y_j P(Y = y_j)$
  - The variance is $\sigma^2 = Var(Y) = E((Y-\mu)^2) = E[Y^2] - (E[Y])^2$ where $E(Y^2) = \sum_{j=1}^{m} y_j^2 P(Y = y_j)$

Gaussian/Normal distribution

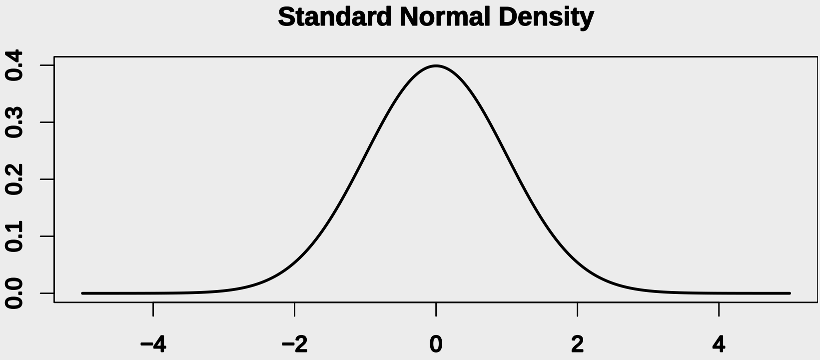
- Let Y be a Normal(μ, σ^2) random variable. It has the density

$$\phi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2} \right\}$$

- Mean / Expected value / Expectation is μ
- Variance is σ^2
- Standard Normal distribution has $\mu = 0$ and $\sigma^2 = 1$

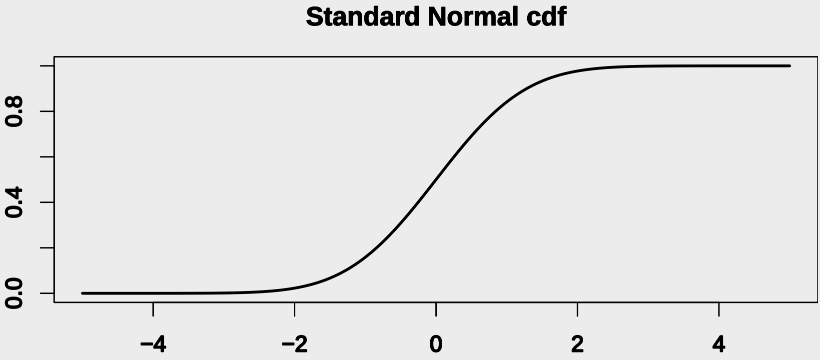
Normal density

```
> y = seq(-5,5,0.01)
> mu = 0
> sigma2 = 1
> plot(y,dnorm(y,mu,sqrt(sigma2)),type="l",lwd=2,main="Standard Normal Density")
```



Normal CDF

```
> y = seq(-5,5,0.01)
> mu = 0
> sigma2 = 1
> plot(y,pnorm(y,mu,sqrt(sigma2)),type="l",lwd=2,main="Standard Normal cdf")
```



Numbers from the standard normal table

Standard Normal Probabilities

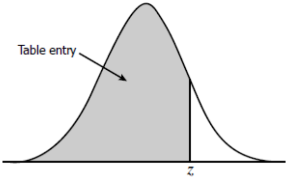


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

Quick Review: Standard Normal Table

Numbers from the standard normal table

1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Quick Review: Using R

Numbers from the standard normal table

```
> qnorm(1-0.05/2,0,1)

[1] 1.959964
```

qnorm is the inverse function of cdf, known as the quantile function. Here Z is a standard normal random variable

$$P(Z \leq 1.959964) = 1 - 0.05/2 = 0.975$$

```
> pnorm(1.959964,0,1)

[1] 0.975
```

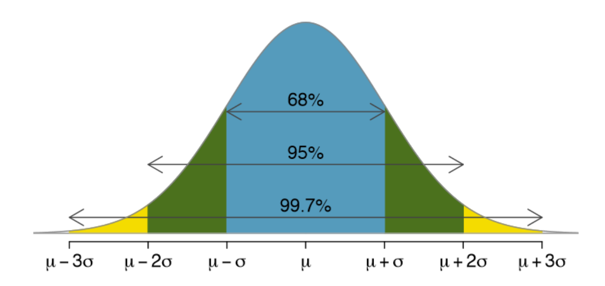
pnorm is the cdf. We have

$$\begin{aligned} P(Z > 1.959964) &= 0.05/2 \\ P(-1.959964 \leq Z \leq 1.959964) &= 1 - 0.05 = 95\% \end{aligned}$$

When we draw a standard normal random number, 95% chance that the number would be in the interval $(-1.959964, 1.959964)$ or $(-1.96, 1.96)$.

Quick Review: Normal probability

In general, if Y is a Normal(μ, σ^2) random variable



By transformation, $Z = \frac{Y-\mu}{\sigma}$ is a standard normal random variable:

$$\begin{aligned}
P(-1.96 \leq Z \leq 1.96) &= 1 - 0.05 = 95\% \\
P(-2 \leq Z \leq 2) &\approx 1 - 0.05 = 95\% \\
P(-2 \leq \frac{Y-\mu}{\sigma} \leq 2) &\approx 1 - 0.05 = 95\% \\
P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) &\approx 1 - 0.05 = 95\%
\end{aligned}$$

Quick Review: Law of Large Numbers (LLN)

Law of large numbers

- Let $Y_1, \ldots, Y_n$ be a random sample of size n
- All $Y_1, \ldots, Y_n$ are independent
- They all have the same mean (or expected value) and variance

$$\begin{aligned} \mu &= E[Y_1] = \cdots = E[Y_n] \\ \sigma^2 &= Var(Y_1) = \cdots = Var(Y_n) \end{aligned}$$

- The average is $\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j$
- $\bar{Y}$ is a random variable and μ is a constant
- When $n \to \infty$, then $\bar{Y} \to \mu$

Quick Review: LLN

Example: Sample from Normal$(1, 1)$ distribution of different sizes, 10, 100, 1000, 10000, we compute the average of the sample (noted that the population mean is 1)

```
> mean(rnorm(10,1,1))

[1] 1.074626

> mean(rnorm(100,1,1))

[1] 1.043359

> mean(rnorm(1000,1,1))

[1] 1.020245

> mean(rnorm(10000,1,1))

[1] 0.9981419
```

Central Limit Theorem

- Let $Y_1, \ldots, Y_n$ be a random sample of size n
- All $Y_1, \ldots, Y_n$ are independent
- They all have the same mean (or expected value) and variance

$$\begin{aligned} \mu &= E[Y_1] = \cdots = E[Y_n] \\ \sigma^2 &= Var(Y_1) = \cdots = Var(Y_n) \end{aligned}$$

- The average is $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$
- $\bar{Y}$ is a random variable and both μ and σ^2 are constants
- When $n \to \infty$, then the distribution of $\frac{\bar{Y}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \to$ the distribution of Z where Z is a standard normal random variable. Usually we write $\frac{\bar{Y}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \ \to \ Z \sim \ \text{Normal}(0,1)$
- Both $\frac{\bar{Y}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$ and Z are random variables

Quick Review

Example: Sample $Y_1, \ldots, Y_n$ from (continuous) Uniform$(0, 1)$ distribution of size n. The mean and variance are $\mu = 1/2$ and $\sigma^2 = 1/12$. We compute the average $\bar{Y}$ and compute the standardized value, that is $\frac{\bar{Y}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$.

Here we take $n = 10$.

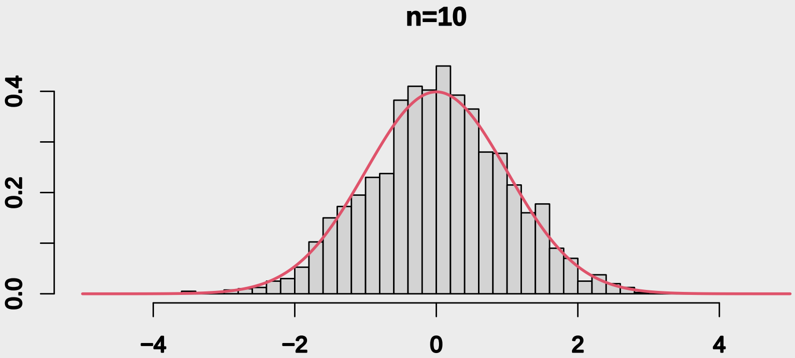
```
> (mean(runif(10,0,1))-1/2)/sqrt(1/12/10)

[1] -0.3924983
```

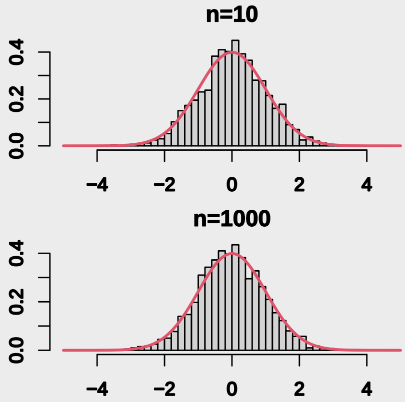
We can't plot it to make comparsion with standard normal density since we have only one value. We need replications.

Example: We repeat this 2000 times so that we can plot the histogram

```
> Standardized.Y = rep(0,2000)
> for (j in 1:2000) {
+   Standardized.Y[j] = (mean(runif(10,0,1))-1/2)/sqrt(1/12/10)
+   }
> hist(Standardized.Y,freq=FALSE,main="n=10",breaks=seq(-5,5,0.2))
> lines(seq(-5,5,0.01),dnorm(seq(-5,5,0.01),0,1),type="l",lwd=2,col=2)
```



Example: We also take $n = 100$, $n = 1000$, and $n = 10000$



Quick Review: t-distribution

Use of t-distribution
Consider the case that σ^2 is unknown. Suppose that $Y_1, \ldots, Y_n$ is a random sample from Normal(μ, σ^2) where both (μ, σ^2) are unknown. We need to estimate σ^2 using the sample variance

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \bar{Y})^2$$

The following statistic has t-distribution with degree of freedom $n-1$

$$T = \frac{\bar{Y} - \mu}{\sqrt{\frac{S^2}{n}}}$$

Note that a statistic is a numerical quantity calculated from data.

Quick Review: t-distribution

Example: Let Y_1, Y_2, Y_3 be the random sample from the Normal$(1, 1)$ distribution. The mean and variance are $\mu = 1$ and $\sigma^2 = 1$. We compute the average $\bar{Y}$ and $\frac{\bar{Y}-\mu}{\sqrt{\frac{S^2}{3}}}$.

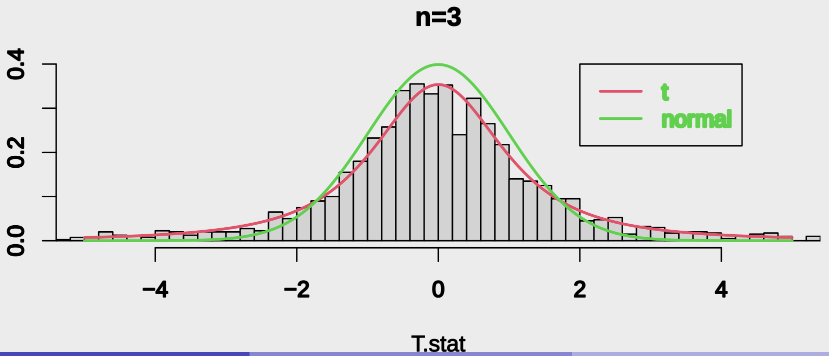
```
> Y = rnorm(3,1,1)
> (mean(Y)-1)/sqrt(var(Y)/3)

[1] -0.3815971
```

Again, we can't plot it to make comparison with a t-density since we have only one value. We need replications.
Note that `rnorm` generates random samples from Normal distribution. The first argument is the size, second argument is the mean and the third argument is the standard error. Entering `?rnorm` in the R console would give you the details. `var(Y)` calculates the sample variance of the data vector `Y`.

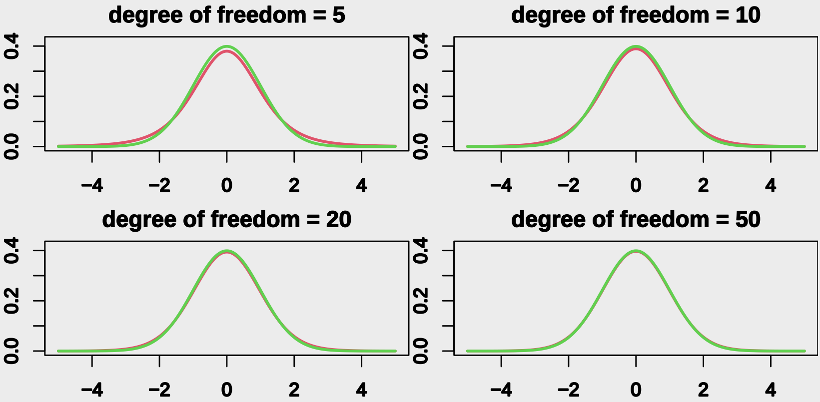
Example: We repeat this 2000 times so that we can plot the histogram

```
> T.stat = rep(0,2000)
> for (j in 1:2000) {
+   Y = rnorm(3,1,1)
+   T.stat[j] = (mean(Y)-1)/sqrt(var(Y)/3)
+   }
> hist(T.stat,freq=FALSE,main="n=3",breaks=2500,xlim=c(-5,5),ylim=c(0,0.4))
> lines(seq(-5,5,0.01),dt(seq(-5,5,0.01),2),type="l",lwd=2,col=2)
```



Quick Review: t-distribution

Comparisons (Normal vs t)



Green: normal; Red: t

When degree of freedom $\to \infty$, t $\to$ normal

Numbers from the t-Table

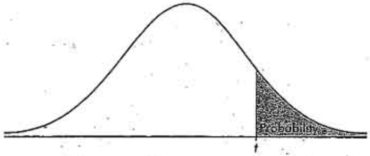


TABLE B: t-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073

Numbers from the t-Table

```
> qt(1-0.05/2,11)

[1] 2.200985
```

T_{11} is a random variable of the t-distribution with degree of freedom 11

$$P(T_{11} \leq 2.200985) \quad = \quad 1 - 0.05/2 = 0.975$$

```
> pt(2.200985,11)

[1] 0.975
```

We have

$$\begin{aligned} P(T_{11} > 2.200985) &= 0.05/2 \\ P(-2.200985 \leq T_{11} \leq 2.200985) &= 1 - 0.05 = 95\% \end{aligned}$$

The random number would be in the interval $(-2.201, 2.201)$ with 95% chance.

Aim 2.4 Quick Review 4: Simple Matrix Algebra

- Two matrices **A** and **B** are equal if they are of the same order and each pair of corresponding elements are equal, i.e., $a_{ij} = b_{ij}$ for $i = 1, 2 \ldots, m$ and $j = 1, 2, \ldots, n$.

```
> A = matrix(c(1,2,3,4),2,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    1    2
[2,]    3    4
> B = matrix(c(1,2,3,4),2,2,byrow=TRUE)
> A==B
     [,1] [,2]
[1,] TRUE TRUE
[2,] TRUE TRUE
> all(A==B)
[1] TRUE
> C = matrix(c(1,2,3,400),2,2,byrow=TRUE)
> all(A==C)
[1] FALSE
```

Simple Matrix Algebra

- A matrix with all elements 0 is denoted by $\mathbf{0}$, i.e.,if $a_{ij} = 0$ for $i = 1, 2 \ldots, m$ and $j = 1, 2, \ldots, n$, then $\mathbf{A} = \mathbf{0}$.

```
> A = matrix(c(0,0,0,0,0,0),3,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    0    0
[2,]    0    0
[3,]    0    0
```

- A square matrix $\mathbf{A}$ with all elements not on the diagonal equal to 0 is a diagonal matrix, i.e.,if $a_{ij} = 0$ for all $i \neq j$ (and $a_{ii} \neq 0$ for at least one i).

```
> A = matrix(c(1,0,0,0,2,0,0,0,3),3,3,byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    2    0
[3,]    0    0    3
```

Simple Matrix Algebra

- A diagonal matrix with all diagonal elements 1 (and all others not on the diagonal 0) is denoted by $\mathbf{I}_n$ or simply $\mathbf{I}$, i.e.,if $a_{ii} = 1$ for $i = 1, 2 \ldots, n$, $a_{ij} = 0$ for for all $i \neq j$ for $i = 1, 2 \ldots, n$, then $\mathbf{A} = \mathbf{I}_n = \mathbf{I}$. It is referred to as the identity matrix

$$\mathbf{I}_n = \mathbf{I} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{n \times n}$$

Simple Matrix Algebra

- identity matrices, $\mathbf{I}_3$:

```
> I = matrix(c(1,0,0,0,1,0,0,0,1),3,3,byrow=TRUE)
> I
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

$\mathbf{I}_4$:

```
> I = matrix(c(1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1),4,4,byrow=TRUE)
> I
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
```

- If $\mathbf{A}$ is $n \times m$ and $\mathbf{I}_n$ is $n \times n$ identity matrix, then $\mathbf{A} = \mathbf{I}_n\mathbf{A}$.
- If $\mathbf{A}$ is $n \times m$ and $\mathbf{I}_m$ is $m \times m$ identity matrix, then $\mathbf{A} = \mathbf{A}\mathbf{I}_m$.

```
> I3 = matrix(c(1,0,0,0,1,0,0,0,1),3,3,byrow=TRUE)
> I2 = matrix(c(1,0,0,1),2,2,byrow=TRUE)
> A = matrix(c(3,3,4,5,6,7),3,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    3    3
[2,]    4    5
[3,]    6    7
> A%*%I2
     [,1] [,2]
[1,]    3    3
[2,]    4    5
[3,]    6    7
> I3%*%A
     [,1] [,2]
[1,]    3    3
[2,]    4    5
[3,]    6    7
```

Simple Matrix Algebra

- If $\mathbf{A}$ is $n \times n$ and $\mathbf{I}_n$ is $n \times n$ identity matrix, then $\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{I}_n\mathbf{A}$.

```
> I = matrix(c(1,0,0,0,1,0,0,0,1),3,3,byrow=TRUE)
> A = matrix(c(3,3,4,5,6,7,3,6,1),3,3,byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    3    3    4
[2,]    5    6    7
[3,]    3    6    1
> A%*%I
     [,1] [,2] [,3]
[1,]    3    3    4
[2,]    5    6    7
[3,]    3    6    1
> I%*%A
     [,1] [,2] [,3]
[1,]    3    3    4
[2,]    5    6    7
[3,]    3    6    1
```

Simple Matrix Algebra

- If $\mathbf{A}$ is a square matrix then diag($\mathbf{A}$) is the (column) vector of the diagonal elements of $\mathbf{A}$, i.e.,the vector (a_{ii}).

$$\mathbf{A} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}}_{n \times n}$$

then

$$\text{diag}(\mathbf{A}) = \underbrace{\begin{bmatrix} a_{11} \\ a_{22} \\ \vdots \\ a_{nn} \end{bmatrix}}_{n \times 1}$$

Simple Matrix Algebra

- square matrix

```
> A = matrix(c(1,2,3,4,1,4,3,2,1),3,3,byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    1    4
[3,]    3    2    1
> diag(A)
[1] 1 1 1
> matrix(diag(A),3,1,byrow=TRUE)
     [,1]
[1,]    1
[2,]    1
[3,]    1
```

Simple Matrix Algebra

- If **a** is a vector, diag(**a**) is the diagonal matrix with the elements of **a** along the diagonal and 0s elsewhere. Let

$$\mathbf{a} = \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}}_{n \times 1}$$

then

$$\text{diag}(\mathbf{a}) = \underbrace{\begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}}_{n \times n}$$

Simple Matrix Algebra

- **diag**

```
> a = matrix(c(1,2,3),3,1,byrow=TRUE)
> a
     [,1]
[1,]    1
[2,]    2
[3,]    3
> diag(c(a))
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    2    0
[3,]    0    0    3
```

Simple Matrix Algebra

- $\mathrm{diag}(\mathrm{diag}(\mathbf{X}))$ is a square matrix formed by setting all off-diagonal elements of $\mathbf{A}$ to 0.

$$\mathrm{diag}(\mathrm{diag}(\mathbf{A})) = \underbrace{\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}}_{n \times n}$$

Some texts will call $\mathrm{diag}(\mathbf{A})$ this matrix but the form `diag(diag(A))` here conforms with `R` syntax.

Simple Matrix Algebra

- double `diag`

```
> A = matrix(c(1,2,3,4,1,4,3,2,1),3,3,byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    1    4
[3,]    3    2    1
> diag(diag(A))
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Simple Matrix Algebra

- The trace of a square matrix is the sum of all the diagonal elements of the matrix, i.e., $\text{tr}(\mathbf{A}) = \text{trace}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$. Note that $\text{tr}(\mathbf{I}_n) = n$.

```
> A = matrix(c(6,2,3,4,2,4,3,2,7),3,3,byrow=TRUE)
> A

     [,1] [,2] [,3]
[1,]    6    2    3
[2,]    4    2    4
[3,]    3    2    7

> sum(diag(A))

[1] 15

> B = matrix(c(1,0,0,0,1,0,0,0,1),3,3,byrow=TRUE)
> B

     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1

> sum(diag(B))

[1] 3
```

Simple Matrix Algebra

- Addition and subtraction of matrices of the same order are performed element by element (just as with vectors):

$$\mathbf{A}+\mathbf{B} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{m\times n} + \underbrace{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}}_{m\times n}$$

$$= \underbrace{\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}}_{m\times n}$$

Simple Matrix Algebra

- Addition and Subtraction

```
> A = matrix(c(6,2,3,4,2,4),3,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    6    2
[2,]    3    4
[3,]    2    4
> B = matrix(c(1,4,2,4,2,3),3,2,byrow=TRUE)
> B
     [,1] [,2]
[1,]    1    4
[2,]    2    4
[3,]    2    3
> A+B
     [,1] [,2]
[1,]    7    6
[2,]    5    8
[3,]    4    7
```

Simple Matrix Algebra

- If **A** and **B** are matrices then we can multiply **A** by **B** (to get **AB**) only if the number of columns of **A** equals the number of rows of **B**:

$$
\mathbf{AB} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nq} \end{bmatrix}}_{n\times q}
$$

$$
= \underbrace{\begin{bmatrix} \sum_{k=1}^n a_{1k}b_{k1} & \sum_{k=1}^n a_{1k}b_{k2} & \cdots & \sum_{k=1}^n a_{1k}b_{kq} \\ \sum_{k=1}^n a_{2k}b_{k1} & \sum_{k=1}^n a_{2k}b_{k2} & \cdots & \sum_{k=1}^n a_{2k}b_{kq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^n a_{mk}b_{k1} & \sum_{k=1}^n a_{mk}b_{k2} & \cdots & \sum_{k=1}^n a_{mk}b_{kq} \end{bmatrix}}_{m\times q}
$$

the element of **AB** at (ij) cell is given by $\sum_{k=1}^n a_{ik}b_{kj}$

Simple Matrix Algebra

- Multiplication

```
> A = matrix(c(1,2,3,4,1,4),3,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    1    4
> B = matrix(c(1,2,1,4,3,4),2,3,byrow=TRUE)
> B
     [,1] [,2] [,3]
[1,]    1    2    1
[2,]    4    3    4
> A%*%B
     [,1] [,2] [,3]
[1,]    9    8    9
[2,]   19   18   19
[3,]   17   14   17
> B%*%A
     [,1] [,2]
[1,]    8   14
[2,]   17   36
```

Simple Matrix Algebra

- If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times q$, then $(\mathbf{AB})'$ is $q \times m$ and $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

```
> A = matrix(c(6,2,3,4,2,4),3,2,byrow=TRUE)
> B = matrix(c(2,5,1,3,5,3,1,4),2,4,byrow=TRUE)
> t(A%*%B)

     [,1] [,2] [,3]
[1,]   22   26   24
[2,]   36   27   22
[3,]    8    7    6
[4,]   26   25   22

> t(B)%*%t(A)

     [,1] [,2] [,3]
[1,]   22   26   24
[2,]   36   27   22
[3,]    8    7    6
[4,]   26   25   22
```

Simple Matrix Algebra

- If $\mathbf{A}$ is $m \times n$ then the products $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ are both defined resulting in $m \times m$ and $n \times n$ square matrices respectively. Both are symmetric because, for example, $(\mathbf{AA}')' = \mathbf{AA}'$

```
> A = matrix(c(6,2,3,4,2,4),3,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    6    2
[2,]    3    4
[3,]    2    4
> t(A)%*%A
     [,1] [,2]
[1,]   49   32
[2,]   32   36
> t(t(A)%*%A)
     [,1] [,2]
[1,]   49   32
[2,]   32   36
```

- If both **A** and **B** are $m \times m$ then $\text{tr}(\mathbf{A}+\mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$.

```
> A = matrix(c(6,2,1,3,4,2,4,5,2),3,3,byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    6    2    1
[2,]    3    4    2
[3,]    4    5    2
> B = matrix(c(4,2,4,5,6,2,3,2,1),3,3,byrow=TRUE)
> B
     [,1] [,2] [,3]
[1,]    4    2    4
[2,]    5    6    2
[3,]    3    2    1
> sum(diag(A+B))
[1] 23
> sum(diag(A))+sum(diag(B))
[1] 23
```

- If $\mathbf{A}$ and $\mathbf{B}$ are both $m \times n$ and then $\text{tr}(\mathbf{A}'\mathbf{B}) = \text{tr}(\mathbf{A}\mathbf{B}') = \text{tr}(\mathbf{B}'\mathbf{A}) = \text{tr}(\mathbf{B}\mathbf{A}')$.

```
> A = matrix(c(6,2,3,4,2,4),3,2,byrow=TRUE)
> B = matrix(c(4,2,4,6,2,3),3,2,byrow=TRUE)
> sum(diag(t(A)%*%B))
[1] 80
> sum(diag(A%*%t(B)))
[1] 80
> sum(diag(t(B)%*%A))
[1] 80
> sum(diag(B%*%t(A)))
[1] 80
```

Simple Matrix Algebra

- If $\mathbf{A}$ is $m \times n$ then $\text{tr}(\mathbf{A}\mathbf{A}') = \text{tr}(\mathbf{A}'\mathbf{A})$.

```
> A = matrix(c(6,2,3,4,2,4),3,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    6    2
[2,]    3    4
[3,]    2    4
> sum(diag(t(A)%*%A))
[1] 85
> sum(diag(A%*%t(A)))
[1] 85
```

Simple Matrix Algebra

- If the columns of $\mathbf{A}$ are $\mathbf{a}_j$ for $j = 1, \ldots, n$

$$\mathbf{A} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{m \times n} = \underbrace{\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}}_{m \times n}$$

then $\mathbf{A}\mathbf{A}' = \sum_{j=1}^{n} \mathbf{a}_j \mathbf{a}_j'$. Moreover,

$$\mathbf{A}' = \underbrace{\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}}_{n \times m} = \underbrace{\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_n' \end{bmatrix}}_{n \times m}$$

$\mathbf{A}'\mathbf{A}$ is known as the cross-product of $\mathbf{A}$.

Simple Matrix Algebra

- Columns of a matrix

```
> A = matrix(c(6,2,3,4,2,4),3,2,byrow=TRUE)
> A%*%t(A)
     [,1] [,2] [,3]
[1,]   40   26   20
[2,]   26   25   22
[3,]   20   22   20
> a1 = A[,1,drop=FALSE]
> a2 = A[,2,drop=FALSE]
> a1%*%t(a1)+a2%*%t(a2)
     [,1] [,2] [,3]
[1,]   40   26   20
[2,]   26   25   22
[3,]   20   22   20
```

Simple Matrix Algebra

- If $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{B}$ is also an $n \times n$ matrix such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$, the $n \times n$ identity matrix, then $\mathbf{B}$ is the inverse of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.

```
> A = matrix(c(6,4,4,6),2,2,byrow=TRUE)
> A

     [,1] [,2]
[1,]    6    4
[2,]    4    6

> iA = solve(A)
> iA

     [,1] [,2]
[1,]  0.3 -0.2
[2,] -0.2  0.3

> A%*%iA

     [,1] [,2]
[1,]    1    0
[2,]    0    1

> iA%*%A

     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

Simple Matrix Algebra

- Solving the following system of equations for (x, y)

$$\begin{aligned} 3x + 2y &= 16 \\ 4x + 7y &= 23 \end{aligned}$$

this can be rewritten as

$$\underbrace{\begin{bmatrix} 3 & 2 \\ 4 & 7 \end{bmatrix}}_{2\times 2} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{2\times 1} = \underbrace{\begin{bmatrix} 16 \\ 23 \end{bmatrix}}_{2\times 1} \Rightarrow \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{2\times 1} = \underbrace{\begin{bmatrix} 3 & 2 \\ 4 & 7 \end{bmatrix}^{-1}}_{2\times 2} \underbrace{\begin{bmatrix} 16 \\ 23 \end{bmatrix}}_{2\times 1}$$

```
> A = matrix(c(3,2,4,7),2,2,byrow=TRUE)
> b = matrix(c(16,23),2,1,byrow=TRUE)
> solve(A)%*%b
          [,1]
[1,] 5.0769231
[2,] 0.3846154

> solve(A,b)
          [,1]
[1,] 5.0769231
[2,] 0.3846154
```

Simple Matrix Algebra

- We are going compute the expected value of sample variance using matrix notation.
- Let $Y_1, \ldots, Y_n$ be independent Normal$(0, \sigma^2)$ random variables.
- Let

$$\mathbf{Y} = \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{n \times 1}, \mathbf{1} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{n \times 1}$$

- The sum of squares is given by

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{Y} = \mathbf{Y}'\mathbf{D}\mathbf{Y}$$

where $\mathbf{D} = \mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$

- The expected value of the sum of squares is given by

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right] &= E[\mathbf{Y}'\mathbf{D}\mathbf{Y}] = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}E[Y_iY_j] \\
&= \sum_{i=1}^{n} d_{ii}E[Y_i^2] + \sum_{i=1}^{n}\sum_{j\neq i} d_{ij}E[Y_i]E[Y_j] \\
&= \sum_{i=1}^{n} d_{ii}E[Y_i^2] = \sigma^2\text{tr}(\mathbf{D})
\end{aligned}
$$

Here

$$
\begin{aligned}
\text{tr}(\mathbf{D}) &= \text{tr}(\mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}') = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}') \\
&= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{1}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_1) = n - 1
\end{aligned}
$$

So it is clear that $E\left[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right] = \sigma^2$