How a little linear algebra can go a long way in the Math Stat course

Randall Pruim

Calvin College

What my students (sort of) know coming in

In theory, my students know

- How to add/subtract vectors; scalar multiplication
- How to represent vectors in Cartesian coordinates (with strong preference for 2d and 3d)
- How to compute a dot product (perhaps matrix multiplication, too)
- The Pythagorean Theorem and how to compute the length (magnitude) of a vector
- $\mathbf{u} \perp \mathbf{v} \iff \mathbf{u} \cdot \mathbf{v} = 0$
    - Perhaps $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}| \cdot |\mathbf{v}| \cos \theta$
- How to compute projections using dot products
- The notion of a span

In practice, these need to be refreshed a bit for some of them, but they do not find this difficult. (I assign a reading and some problems and only discuss difficulties in class.) They get more fluent as we go along.

Additional Linear Algebra they need

To Sum up . . .

To Sum up . . .

1. Students don't need a lot of linear algebra to make use of linear algebra in statistics

To Sum up . . .

1. Students don't need a lot of linear algebra to make use of linear algebra in statistics
2. The review of basic linear algebra in an application area is *good for their linear algebra*

To Sum up . . .

1. Students don't need a lot of linear algebra to make use of linear algebra in statistics
2. The review of basic linear algebra in an application area is *good for their linear algebra*
3. (A little) linear algebra provides an important perspective on statistics

Linear Algebra and Statistics (1)

Summary: The expected values and variances of linear combinations of independent normal random variables are easily computed.

Example: Suppose $X_1 \sim \text{Norm}(\mu_1, \sigma_1^2)$, and $X_2 \sim \text{Norm}(\mu_2, \sigma_2^2)$.

Then

- $\text{E}(aX_1 + bX_2) = a\mu_1 + b\mu_2$,
- $\text{Var}(aX_1 + bX_2) = a^2\sigma_1^2 + b^2\sigma_2^2$

Linear Algebra and Statistics (1)

Summary: The expected values and variances of linear combinations of independent normal random variables are easily computed.

Example: Suppose $X_1 \sim \text{Norm}(\mu_1, \sigma_1^2)$, and $X_2 \sim \text{Norm}(\mu_2, \sigma_2^2)$.

Then

- $\text{E}(aX_1 + bX_2) = a\mu_1 + b\mu_2$,
- $\text{Var}(aX_1 + bX_2) = a^2\sigma_1^2 + b^2\sigma_2^2$

That is,

- $\text{E}(\langle a, b\rangle \cdot \langle X_1, X_2\rangle) = \langle a, b\rangle \cdot \langle \mu_1, \mu_2\rangle$
- $\text{Var}(\langle a, b\rangle \cdot \langle X_1, X_2\rangle) = \langle a^2, b^2\rangle \cdot \langle \sigma_1^2, \sigma_2^2\rangle$

Bonus: If the component distributions are normal, the linear combination is also normal.

Linear algebra provides notation and perspective (and makes it easier to increase dimienesion).

Linear Algebra and Statistics (2)

If $\mathbf{u} \in \mathbb{R}^n$ is a constant vector and $\mathbf{X} = \langle X_1, X_2, \ldots, X_n \rangle$ is a vector of independent random variables with means $\boldsymbol{\mu} = \langle \mu_1, \mu_2, \ldots, \mu_n \rangle$ and standard deviations $\boldsymbol{\sigma} = \langle \sigma_1, \sigma_2, \ldots, \sigma_n \rangle$, then

1. $\mathrm{E}(\mathbf{u} \cdot \mathbf{X}) = \mathbf{u} \cdot \boldsymbol{\mu}$
2. $\mathrm{Var}(\mathbf{u} \cdot \mathbf{X}) = \mathbf{u}^2 \cdot \boldsymbol{\sigma}^2$

Summary: Linear combinations of independent [normal] random variables are [normal] random variables with means and variances that are easily computed.

Linear Algebra and Statistics (2)

If $\mathbf{u} \in \mathbb{R}^n$ is a constant vector and $\mathbf{X} = \langle X_1, X_2, \ldots, X_n \rangle$ is a vector of independent random variables with means $\boldsymbol{\mu} = \langle \mu_1, \mu_2, \ldots, \mu_n \rangle$ and standard deviations $\boldsymbol{\sigma} = \langle \sigma_1, \sigma_2, \ldots, \sigma_n \rangle$, then

1. $\mathrm{E}(\mathbf{u} \cdot \mathbf{X}) = \mathbf{u} \cdot \boldsymbol{\mu}$
2. $\mathrm{Var}(\mathbf{u} \cdot \mathbf{X}) = \mathbf{u}^2 \cdot \boldsymbol{\sigma}^2$

In particular, if $|\mathbf{u}| = 1$, and $\mathbf{X} \overset{\text{iid}}{\sim} \mathrm{Norm}(\mu, \sigma)$, then

1. $\mathrm{E}(\mathbf{u} \cdot \mathbf{X}) = (\mathbf{u} \cdot \mathbf{1})\mu$
2. $\mathrm{Var}(\mathbf{u} \cdot \mathbf{X}) = (\mathbf{u}^2 \cdot \mathbf{1})\sigma^2 = \sigma^2$
3. $\mathbf{u} \cdot \mathbf{X}$ is a normal random variable

And if, in addition, $\mathbf{u} \perp \mathbf{1}$, then

1. $\mathrm{E}(\mathbf{u} \cdot \mathbf{X}) = 0$

Summary: Linear combinations of independent [normal] random variables are [normal] random variables with means and variances that are easily computed. (Note: There are some important special cases.)

Linear Algebra and Statistics (3)

If $\mathbf{u}_1$ and $\mathbf{u}_2$ are constant vectors in $\mathbb{R}^n$, and $\mathbf{X}$ is a vector of n independent random variables, then

$$\mathbf{u}_1 \perp \mathbf{u}_2 \iff \mathbf{u}_1 \cdot \mathbf{X} \text{ and } \mathbf{u}_2 \cdot \mathbf{X} \text{ are independent.}$$

- Full proof requires n-dimensional change of variables (i.e., Jacobian)
- Proof that $\mathbf{u}_1 \cdot \mathbf{X}$ and $\mathbf{u}_2 \cdot \mathbf{X}$ are *uncorrelated* is easy application of covariance lemmas.

Looking at variance

The definition of sample variance:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

Looking at variance through linear algebra

The definition of sample variance through linear algebra eyes:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

Looking at variance through linear algebra

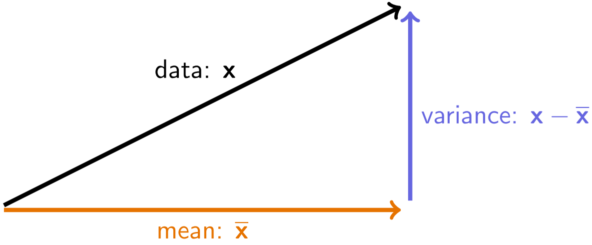
The definition of sample variance through linear algebra eyes:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} = \frac{|\mathbf{x} - \overline{\mathbf{x}}|^2}{n-1}$$

Looking at variance through linear algebra

The definition of sample variance through linear algebra eyes:

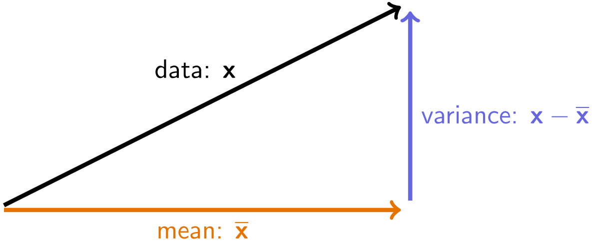
$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} = \frac{|\mathbf{x} - \overline{\mathbf{x}}|^2}{n-1}$$



Looking at variance through linear algebra

The definition of sample variance through linear algebra eyes:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} = \frac{|\mathbf{x} - \overline{\mathbf{x}}|^2}{n-1}$$



$$\overline{\mathbf{x}} = \text{proj}(\mathbf{x} \to \mathbf{1}) \text{ , so } \overline{\mathbf{x}} - \mathbf{x} \perp \mathbf{x}$$

$$(\mathbf{x} - \overline{\mathbf{x}}) \cdot \overline{\mathbf{x}} = \sum (x_i - \overline{x})\overline{x} = \overline{x} \sum (x_i - \overline{x}) = \overline{x}(n\overline{x} - n\overline{x}) = 0$$

Pythagorean decomposition of the variance vector

- $\mathbf{v}_1 = \mathbf{1}$; $\mathbf{u}_1 = \mathbf{1}/\sqrt{n}$
- $\mathbf{u}_2, \ldots, \mathbf{u}_n$ chosen so that
  - Unit length: For all i, $|\mathbf{u}_i| = 1$
  - Orthogonal: Whenever $i \neq j$, $\mathbf{u}_i \perp \mathbf{u}_j$

Pythagorean decomposition of the variance vector

- $\mathbf{v}_1 = \mathbf{1}$; $\mathbf{u}_1 = \mathbf{1}/\sqrt{n}$
- $\mathbf{u}_2, \ldots, \mathbf{u}_n$ chosen so that
  - Unit length: For all i, $|\mathbf{u}_i| = 1$
  - Orthogonal: Whenever $i \neq j$, $\mathbf{u}_i \perp \mathbf{u}_j$

$$\mathbf{X} = \sum_{i=1}^{n} \text{proj}(\mathbf{X} \to \mathbf{u}_i) = \underbrace{\text{proj}(\mathbf{X} \to \mathbf{u}_1)}_{\overline{\mathbf{X}}} + \sum_{i=2}^{n} \text{proj}(\mathbf{X} \to \mathbf{u}_i)$$

$$|\mathbf{X} - \overline{\mathbf{X}}|^2 = \sum_{i=2}^{n} |\text{proj}(\mathbf{X} \to \mathbf{u}_i)|^2 = \sum_{i=2}^{n} \left(\underbrace{\mathbf{u}_i \cdot \mathbf{X}}_{\text{Norm}(0,\sigma)} \right)^2$$

Pythagorean decomposition of the variance vector

- $\mathbf{v}_1 = \mathbf{1}$; $\mathbf{u}_1 = \mathbf{1}/\sqrt{n}$
- $\mathbf{u}_2, \ldots, \mathbf{u}_n$ chosen so that
  - Unit length: For all i, $|\mathbf{u}_i| = 1$
  - Orthogonal: Whenever $i \neq j$, $\mathbf{u}_i \perp \mathbf{u}_j$

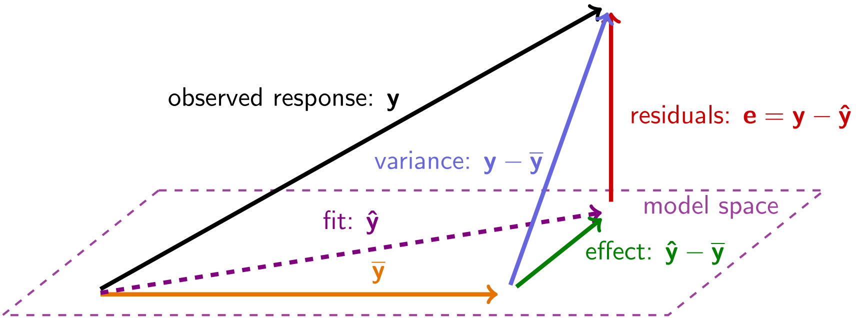
$$\mathbf{X} = \sum_{i=1}^{n} \mathrm{proj}(\mathbf{X} \to \mathbf{u}_i) = \underbrace{\mathrm{proj}(\mathbf{X} \to \mathbf{u}_1)}_{\overline{\mathbf{X}}} + \sum_{i=2}^{n} \mathrm{proj}(\mathbf{X} \to \mathbf{u}_i)$$

$$|\mathbf{X} - \overline{\mathbf{X}}|^2 = \sum_{i=2}^{n} |\mathrm{proj}(\mathbf{X} \to \mathbf{u}_i)|^2 = \sum_{i=2}^{n} \left(\underbrace{\mathbf{u}_i \cdot \mathbf{X}}_{\mathrm{Norm}(0,\sigma)} \right)^2$$

So $\dfrac{(n-1)S^2}{\sigma^2} \sim \mathrm{Chisq}(n-1)$, and $\mathrm{E}(S^2) = \sigma^2$. $\leftarrow n-1$ explained

What does a linear model look like?

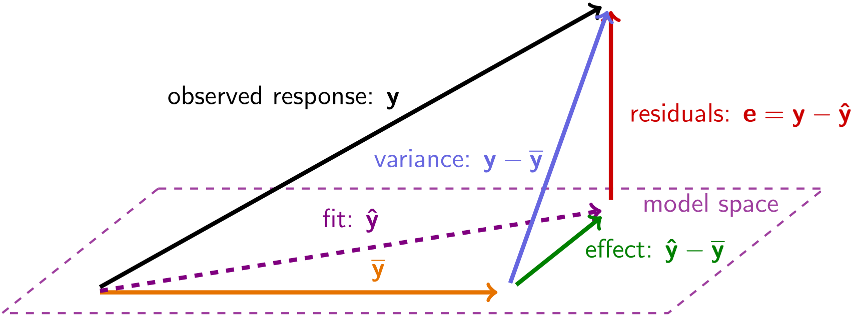
Pythagorean decomposition of $\mathbf{y}$



$\mathbb{R}^n$ spanned by orthogonal vectors $\underbrace{\mathbf{v}_1 = \mathbf{1}, \overbrace{\mathbf{v}_2, \ldots, \mathbf{v}_p}_{}}_{\text{model space}}\!\!\!\!\!\!\!\!\!\!\!\!\!\!\overbrace{\phantom{\mathbf{v}_2, \ldots, \mathbf{v}_p}, \mathbf{v}_{p+1}, \ldots \mathbf{v}_n}^{\text{variance}}$

What does a linear model look like?

Pythagorean decomposition of $\mathbf{y}$



Least Squares: Minimizing $\sum(y_i - \hat{y}_i)^2 = |\mathbf{e}|^2$.

Model Utility Test

H_0: all regression coefficients are 0.

ANOVA (Analysis of Variance) approach: decompose the variance vector

$$\mathbf{y} - \bar{\mathbf{y}} = \hat{\mathbf{y}} - \bar{\mathbf{y}} + \mathbf{e}$$

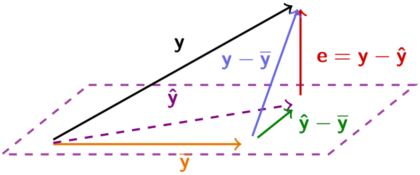
Does $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ do better than random?

If H_0 is true, length of each red and green component is $\text{Norm}(0, \sigma)^2$

The test statistic

$$F = \frac{|\hat{\mathbf{y}} - \bar{\mathbf{y}}|^2 / dfm}{|\mathbf{e}|^2 / dfe} = \frac{MSM}{MSE}$$

should be about 1 when H_0 is true; larger when false.



Model Utility Test

H_0: all regression coefficients are 0.

ANOVA (Analysis of Variance) approach: decompose the variance vector

$$\mathbf{y} - \bar{\mathbf{y}} = \hat{\mathbf{y}} - \bar{\mathbf{y}} + \mathbf{e}$$

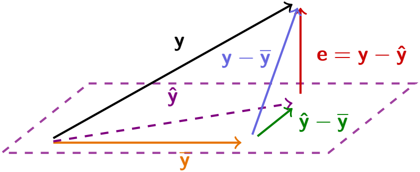
Does $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ do better than random?

If H_0 is true, length of each red and green component is $\text{Norm}(0, \sigma)^2$

The test statistic

$$F = \frac{|\hat{\mathbf{y}} - \bar{\mathbf{y}}|^2 / dfm}{|\mathbf{e}|^2 / dfe} = \frac{MSM}{MSE}$$

should be about 1 when H_0 is true; larger when false.



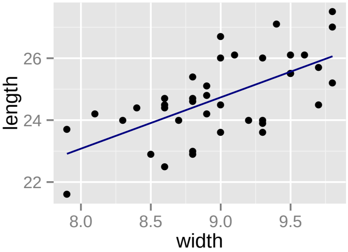
$$R^2 = \frac{SSM}{SSM + SSE} = \cos\theta$$

Model Utility Test

Simple Model

$$\texttt{width} = \beta_0 + \beta_1 \texttt{length} + \varepsilon$$

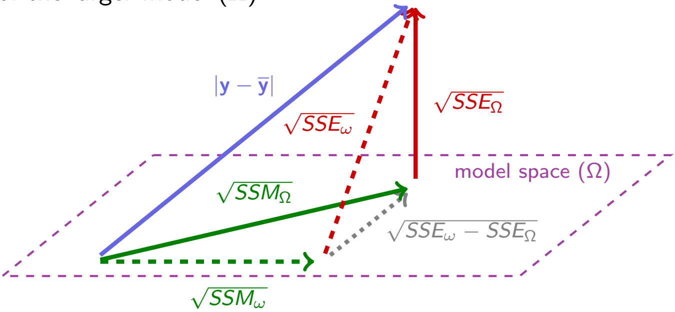
But method works the same way for complex models, too.




```
> model <- lm(width ~ length, KidsFeet)
> anova(model)


Analysis of Variance Table

Response: width
          Df Sum Sq Mean Sq F value  Pr(>F)
length     1   4.06    4.06    25.8 1.1e-05
Residuals 37   5.81    0.16
```

Model Comparison Tests

Nested Models: Model space of smaller model (ω) is a subspace of model space of the larger model (Ω).



H_0: all parameters in the larger model only are 0.

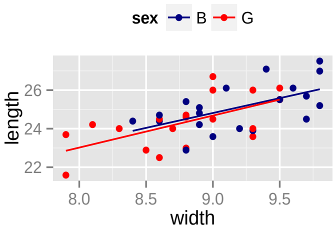
If H_0 is true, then dashed model is correct and dotted gray vector is red.

$$F = \frac{MSM}{MSE} \sim \mathrm{F}(\dim(\Omega) - \dim(\omega), n - \dim(\Omega))$$

Does Sex Matter?

```
> Omega <- lm(width ~ length + sex, data = KidsFeet)
> omega <- lm(width ~ length, data = KidsFeet)
```

```
> anova(omega, Omega)

Analysis of Variance Table

Model 1: width ~ length
Model 2: width ~ length + sex
  Res.Df  RSS Df Sum of Sq     F Pr(>F)
1     37 5.81
2     36 5.33  1     0.479 3.23  0.081
```

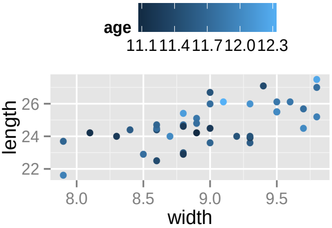



```
> coef(summary(Omega))

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)   3.6412     1.2506   2.912 6.139e-03
length        0.2210     0.0497   4.447 8.015e-05
sexG         -0.2325     0.1293  -1.798 8.055e-02
```

Does Age Matter?

```
> Omega <- lm(width ~ length + age, data = KidsFeet)
> omega <- lm(width ~ length, data = KidsFeet)
```

```
> anova(omega, Omega)


Analysis of Variance Table

Model 1: width ~ length
Model 2: width ~ length + age
  Res.Df  RSS Df Sum of Sq     F Pr(>F)
1     37 5.81
2     36 5.73  1    0.0829 0.52   0.47
```

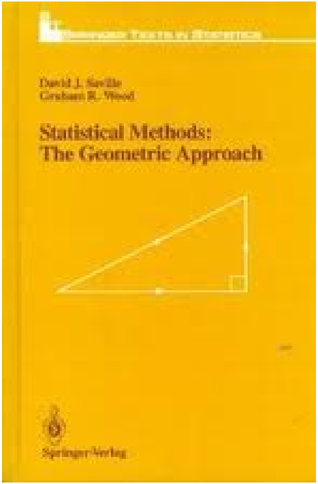


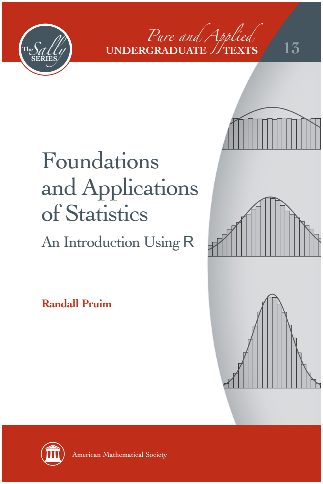

```
> coef(summary(Omega))


             Estimate Std. Error t value  Pr(>|t|)
(Intercept)    1.2855    2.49981  0.5142 6.102e-01
length         0.2331    0.05327  4.3746 9.967e-05
age            0.1667    0.23085  0.7219 4.750e-01
```

To sum up . . .

1. Students don't need a lot of linear algebra to make use of linear algebra in statistics
2. The review of basic linear algebra in an application area is *good for their linear algebra*
3. (A little) linear algebra provides an important perspective on statistics

References


David J. Saville
Graham R. Wood
Statistical Methods:
The Geometric Approach
Springer-Verlag


The Sally SERIES
Pure and Applied
UNDERGRADUATE TEXTS
13
Foundations
and Applications
of Statistics
An Introduction Using R
Randall Pruim
American Mathematical Society