EXAM "CHEAT" SHEET /SUMMARY

Correction Factor

SE without replacement = correction factor × SE with replacement

$$\text{correction factor} = \sqrt{\frac{\text{number of tickets-number of draws}}{\text{number of tickets -1}}}$$

correct

Binomial Theorem

In R: **dbinom(x, n, p)** where,

- x = no. of heads
- n = total number of tosses
- p = probability of event

$$\binom{n}{x} p^x (1-p)^{n-x}$$

SAMPLE

Correlation

- cor
- **r p**

SU

Coefficient

$$CV = \frac{\text{standard deviation}}{}$$

standard deviation

Regression Line

- **lm(y~x)**

SD Line	Regression Line

$(\bar{x}, \bar{y})$ to

$$\frac{SD_y}{SD_x}$$

- **Predi**
- **Predi**

SAMPLE

of small

Standard Deviation

Population

RMS of gaps from the mean $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(gaps)^2}$

```
sd(data)*sqrt((n-1)/n))
```

Residuals

- **RM**

$$\text{RMS of (gaps from the line)} = \sqrt{\text{mean of (gaps)}^2}$$

- **Baseline Prediction**

$$\text{RMS error}_{Pop} = SD_y$$

- **Normal Prediction**

$$SE_{proportion} = \frac{SD_{box}}{\sqrt{\text{sample size}}}$$

Mean Draws from Box

$$EV_{mean} = \frac{EV_{sum}}{\text{number of draws}} = \text{mean}_{box}.$$

aws

Vertical St

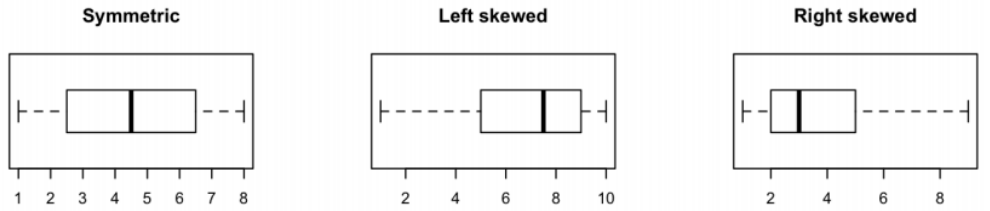
SAMPLE

Chance

- C

$P(\text{Event1}|\text{Event2})$

Addition Rule	events occurs)	P(Event1) + P(Event2)	if mutually exclusive
Multiplication Rule	P(Both events occur)	P(Event1) x P(Event2)	if independent
		P(Event1) x P(Event2, given Event 1)	if dependent



Comparing SD Line and Regression Line

- Cannot interchange variables like the SD Line

Feature	SD Line	Regression Line
Connects	$(\bar{x}, \bar{y})$ to $(\bar{x} + \text{SD}_x, \bar{y} + \text{SD}_y)$	$(\bar{x}, \bar{y})$ to $(\bar{x} + \text{SD}_x, \bar{y} + r\text{SD}_y)$
Slope (b)	$\frac{\text{SD}_y}{\text{SD}_x}$	$r\frac{\text{SD}_y}{\text{SD}_x}$

Graph of averages

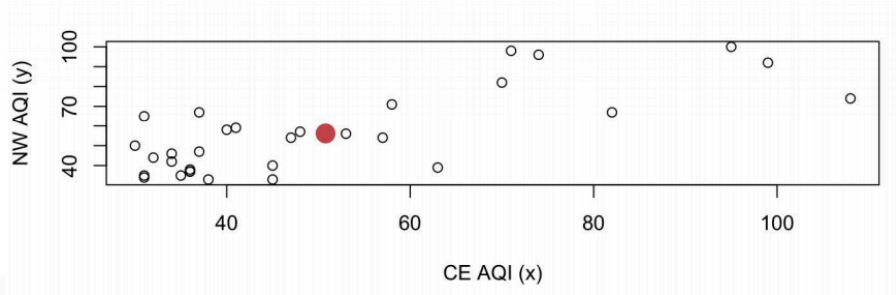
The graph of averages plots the average y for each x.

- The regression line is a smoothed version of the graph of averages.
- If the graph of averages is a smooth line, that line is the regression line.

Predictions

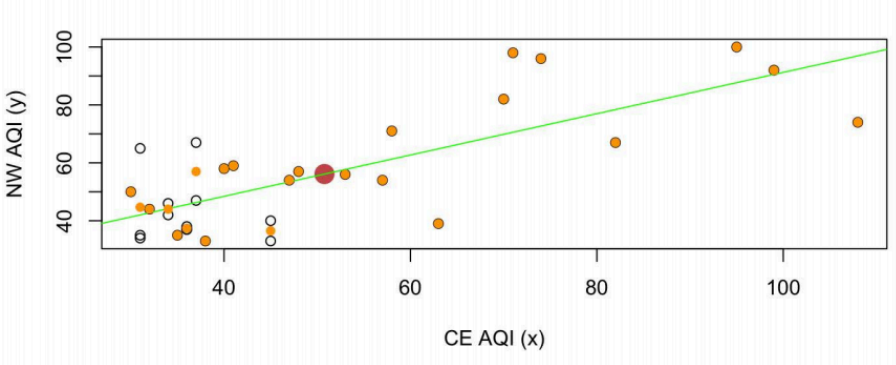
1. Baseline predication

Given a certain value x, a basic prediction of y would be the **average** of y over **all** the x values in the data.



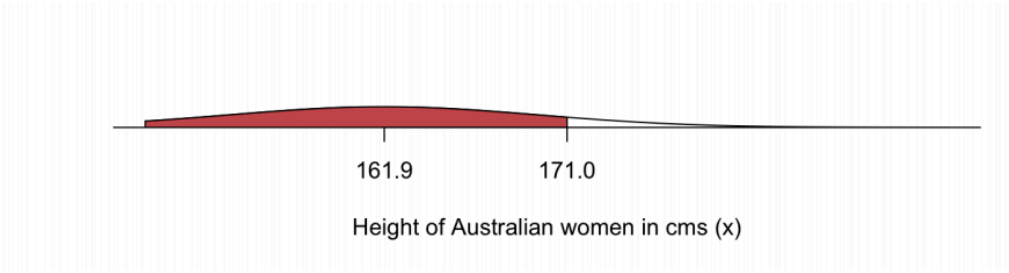
2. Prediction in a strip

- Given a certain value x, a more careful prediction of y would be the average of all the y values in the data corresponding to that x value.
- We use the graph of averages.



Area under a General Normal Curve

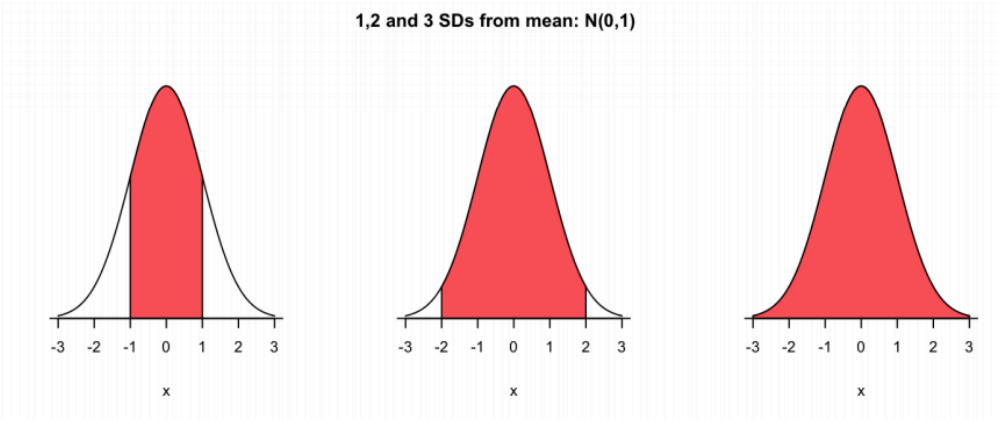
- **Lower tail:** pnorm(x, mean, sd) **Upper tail**: pnorm(x, mean, sd, lower.tail = F)



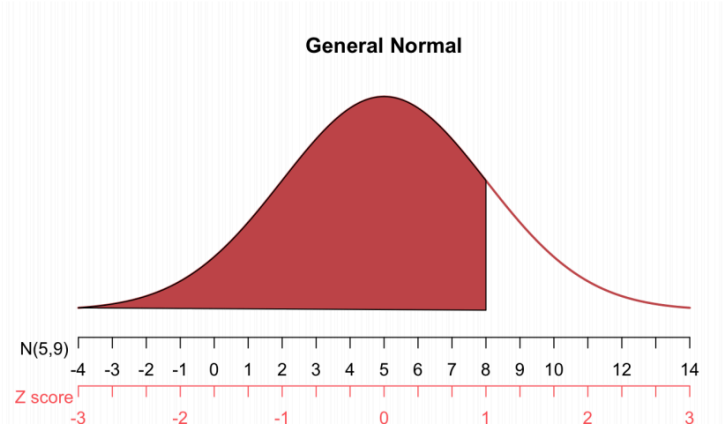

```
pnorm(171, 161.9, 7.7)  #pnorm(x,mean,sd)
```

Special Properties of Normal Curve

1. **All Normal curves satisfy the "68%-95%-99.7% Rule**
   - The area 1 SD out from the mean in both directions is 0.68 (68%).
   - The area 2 SDs out from the mean in both directions is 0.95 (95%).
   - The area 3 SDs out from the mean in both directions is 0.997 (99.7%)



2. **Any General Normal can be rescaled into the Standard Normal**
   - Change to standard units/z score



Here the point = 8.

So the z score is $\frac{8-5}{3} = 1$.