MM104/MM106/BM110
Statistics and Data Presentation

Lecture 5:
Sampling distributions
Distribution of the sample mean

Chris Robertson

DEPARTMENT OF MATHEMATICS & STATISTICS

University of Strathclyde Science

Central Limit Theorem

POPULATION!

- If x random with $\boxed{\text{mean } \mu \text{ and standard deviation } \sigma,}$ after random sampling with n measurements (*sample size*):

CENTRAL LIMIT THEOREM

  - If x normally distributed, sampling distribution of $\bar{x}$ *exactly* normally distributed.
  - If x non-normally distributed, sampling distribution of $\bar{x}$ *approx*. normally distributed.

The larger the sample size, the closer to normality.

Central Limit Theorem

POPULATION!

- If x random with $\boxed{\text{mean } \mu \text{ and standard deviation } \sigma,}$ after random sampling with n measurements (*sample size*):

CENTRAL LIMIT THEOREM

— sampling distribution of $\bar{X}$ has

— mean μ and

— standard deviation $\boxed{\sigma/\sqrt{n}\ .}$

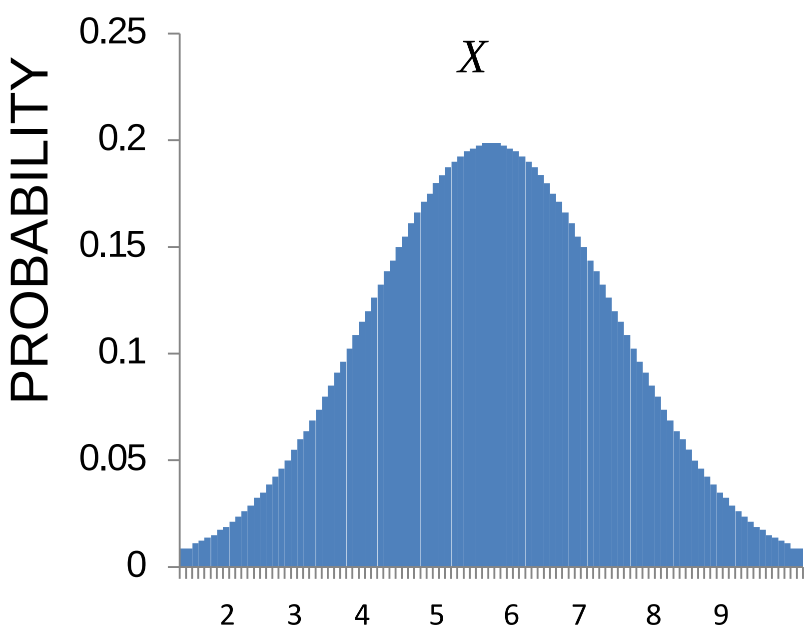
(STANDARD ERROR OF THE SAMPLE MEAN)

Central Limit Theorem

- This is a very important mathematical result in Statistics
- It means that you can use a normal distribution to calculate probabilities associated with sampling distributions even is the original variable is not normally distributed
- We will use it to calculate probabilities about sample means even when the original variable is not normally distributed – such as the mean income when the distribution of income is skew with a long tail to the right (high incomes)
- We will also use it to calculate probabilities about proportions which must lie between 0 and 1 and so cannot really be normally distributed.

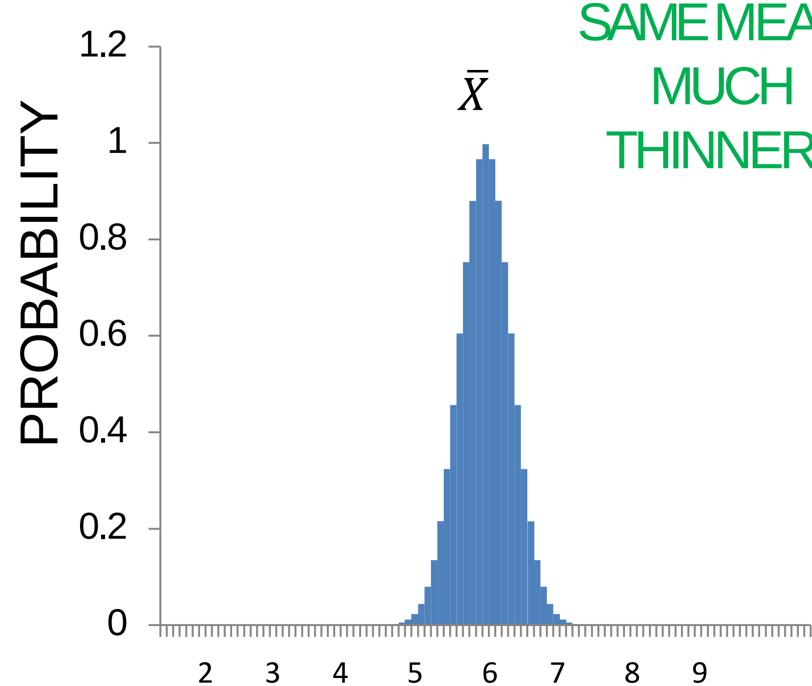
Original Distribution and Sampling distribution

- Regardless of shape for x, in large samples the sampling mean distribution has mean μ and standard error $\sigma/\sqrt{n}$



Mean μ, standard dev. σ

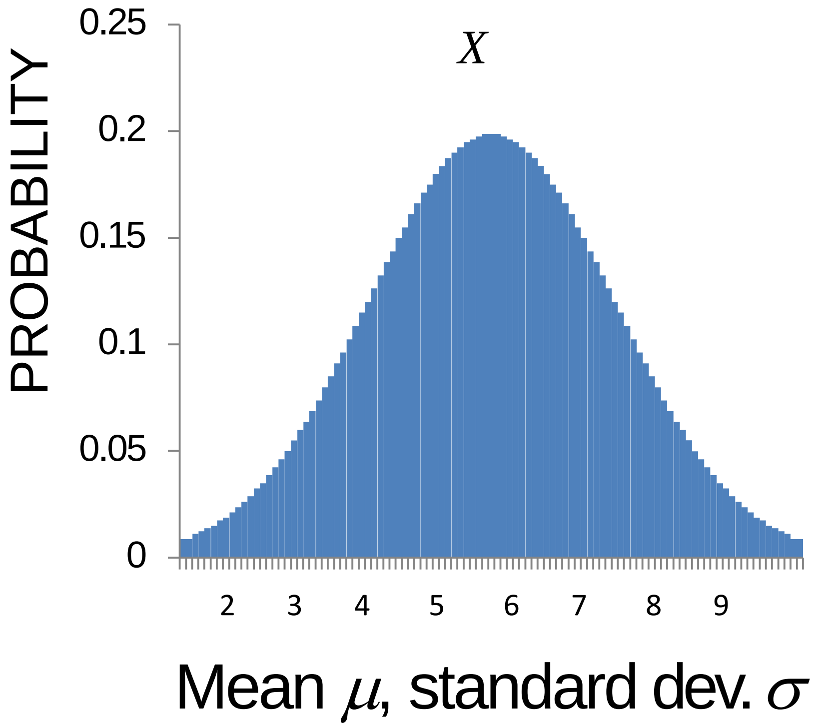
SAME MEAN, MUCH THINNER!!

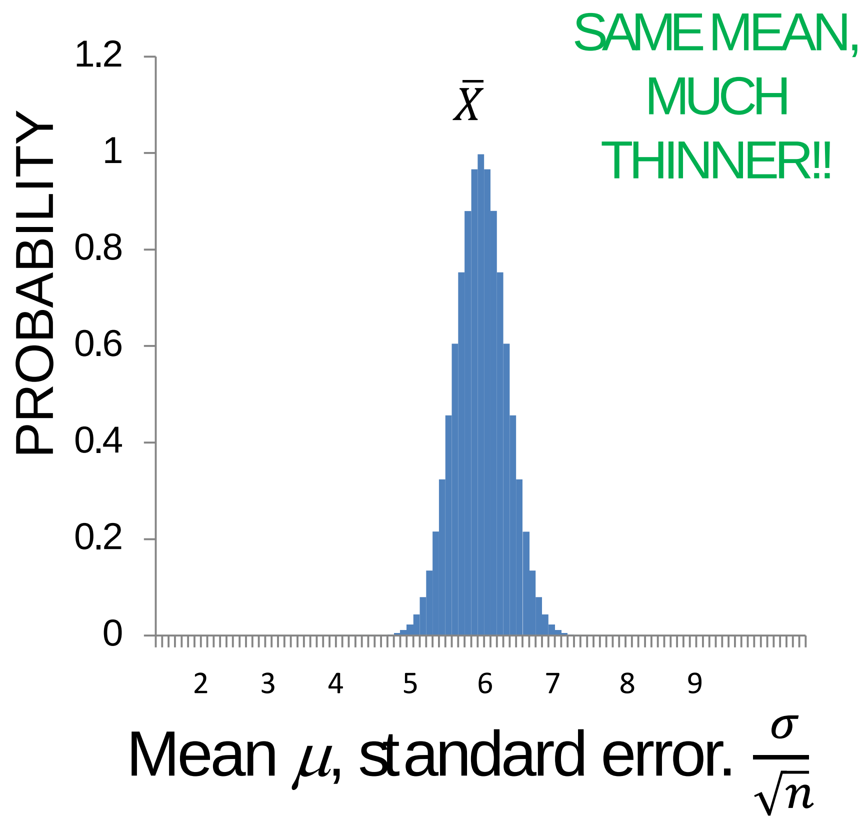


Mean μ, standard error. $\frac{\sigma}{\sqrt{n}}$

Original Distribution and Sampling distribution

Much thinner means that in the sampling distribution of the sample mean more of the probability is closer to the population mean μ than in the distribution of X $\quad P(5 < X < 7) < P(5 < \bar{X} < 7)$

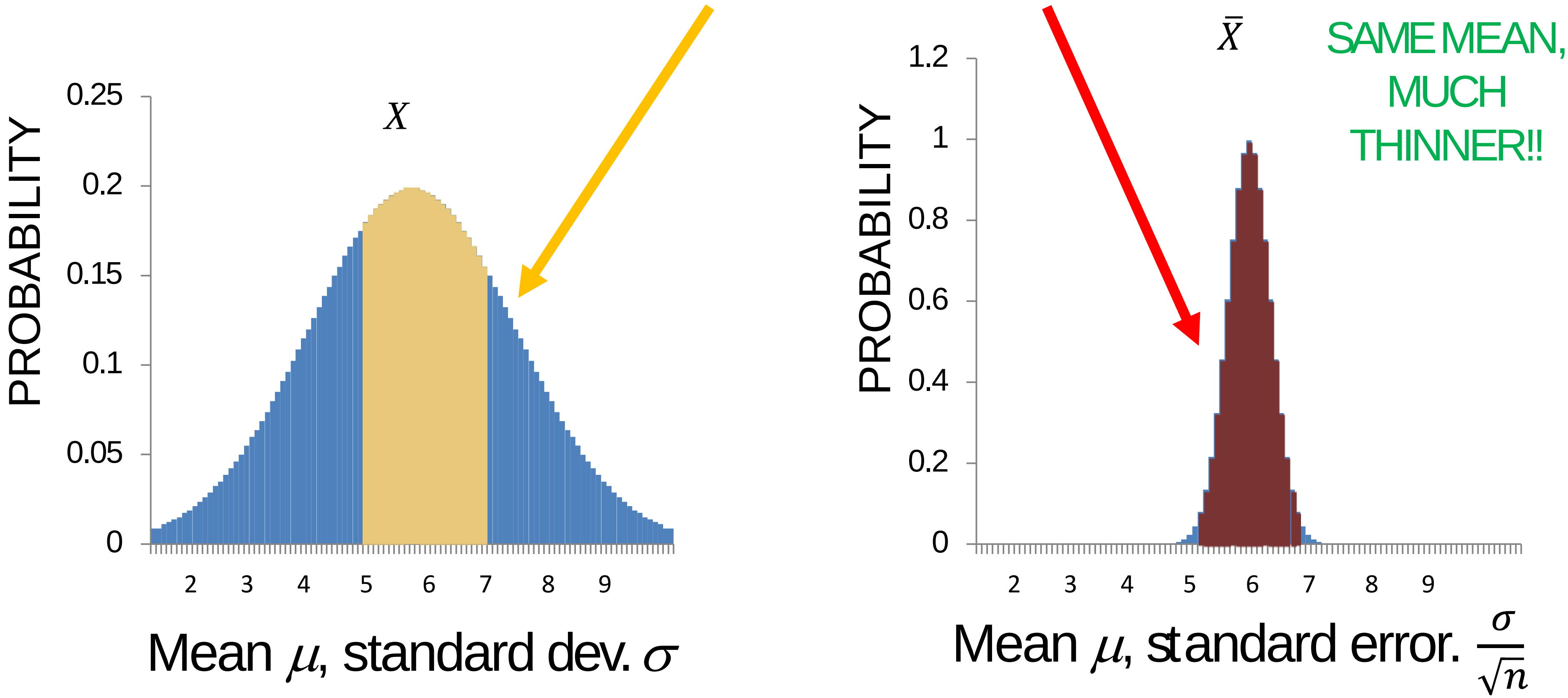




Original Distribution and Sampling distribution

Much thinner means that in the sampling distribution of the sample mean more of the probability is closer to the population mean μ than in the distribution of X

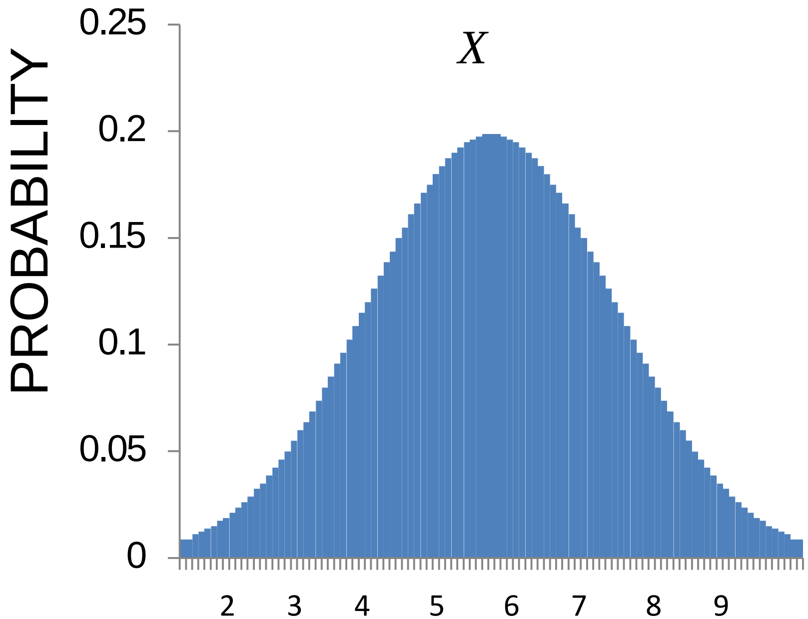
$$P(5 < X < 7) < P(5 < \bar{X} < 7)$$



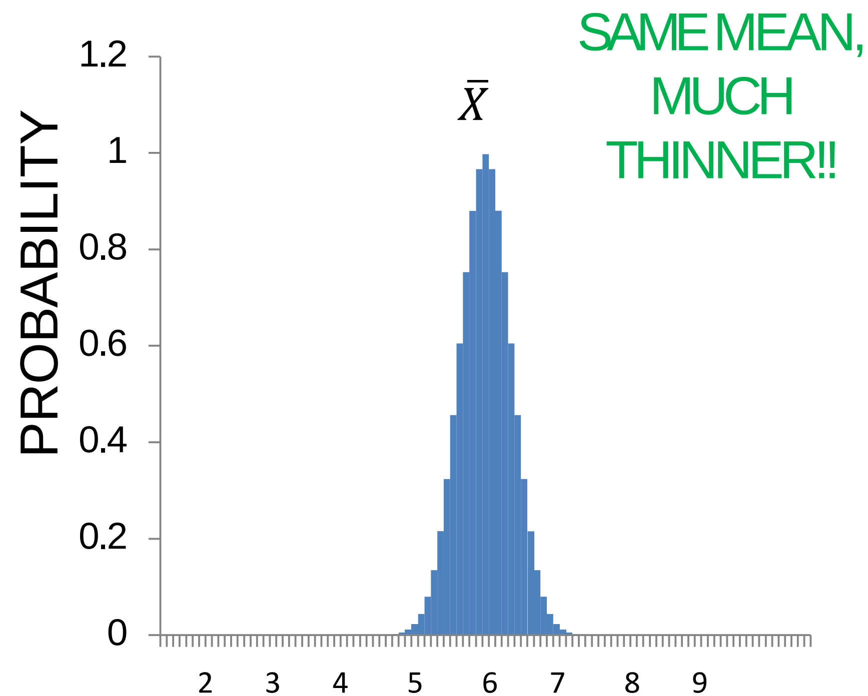
Original Distribution and Sampling distribution

It also means that the probability that the sample mean is far away from the population mean μ is less than in than in the original distribution of *x*

$$P(X > 7) > P(\bar{X} > 7)$$



Mean μ, standard dev. σ

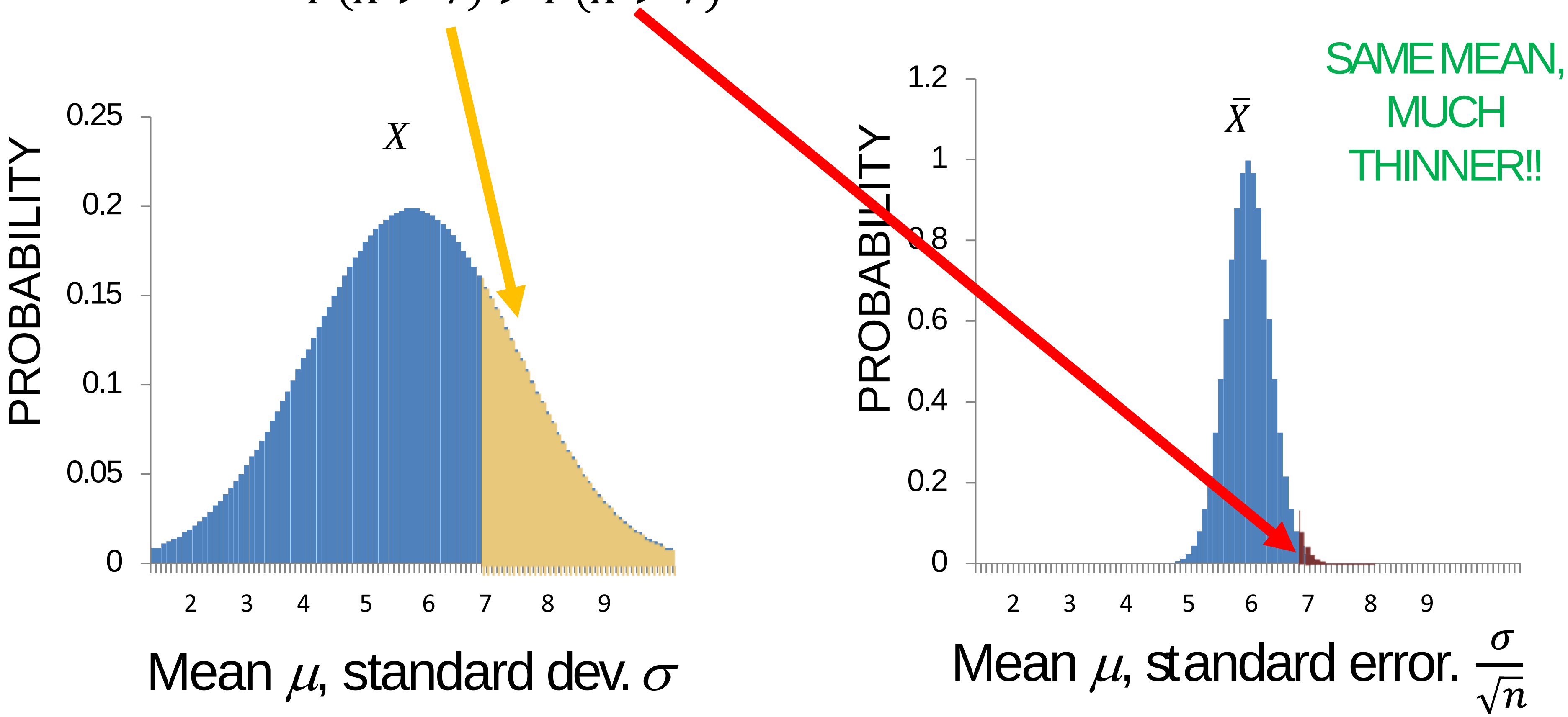


Mean μ, standard error. $\frac{\sigma}{\sqrt{n}}$

Original Distribution and Sampling distribution

It also means that the probability that the sample mean is far away from the population mean μ is less than in than in the original distribution of x

$$P(X > 7) > P(\bar{X} > 7)$$



Sampling distribution – probability calculation

- In a large company, the number of hours that full-time employees work in any week is known to be Normally distributed with $\mu = 30$h and standard deviation $\sigma = 3.9$h. An employee is paid overtime if she/he works more than 31.5 hours per week. What is the probability that a randomly selected employee will be paid overtime in any one week?

$x \equiv$ no. hours worked per week

$$P(x > 31.5)?$$

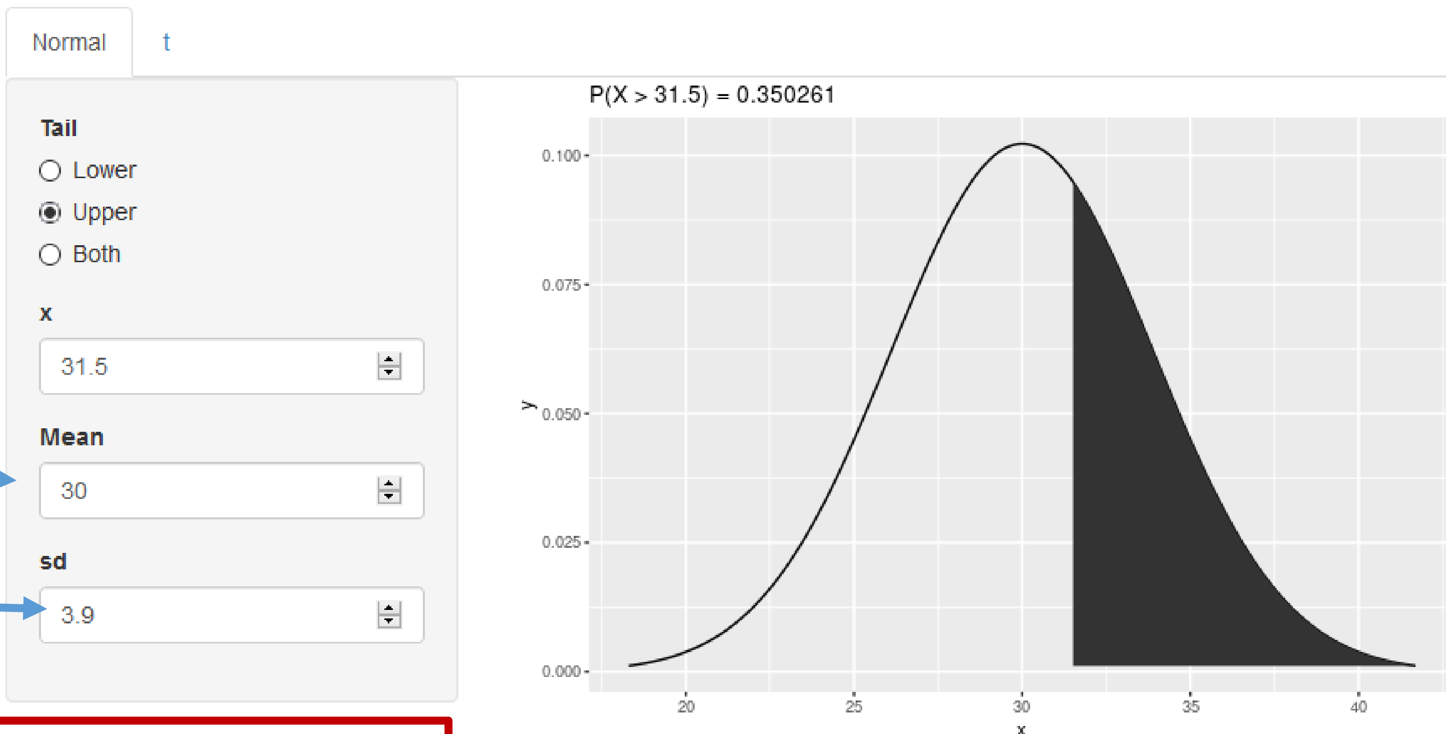
Use the tables of the normal distribution with a mean of 30 and standard deviation of 3.9

Sampling distribution – probability calculation

$x \equiv$ no. hours worked per week

$$P(x > 31.5) = 0.35026$$

Use the tables of the normal distribution with a mean of 30 and standard deviation of 3.9



Fill in the mean and standard deviation

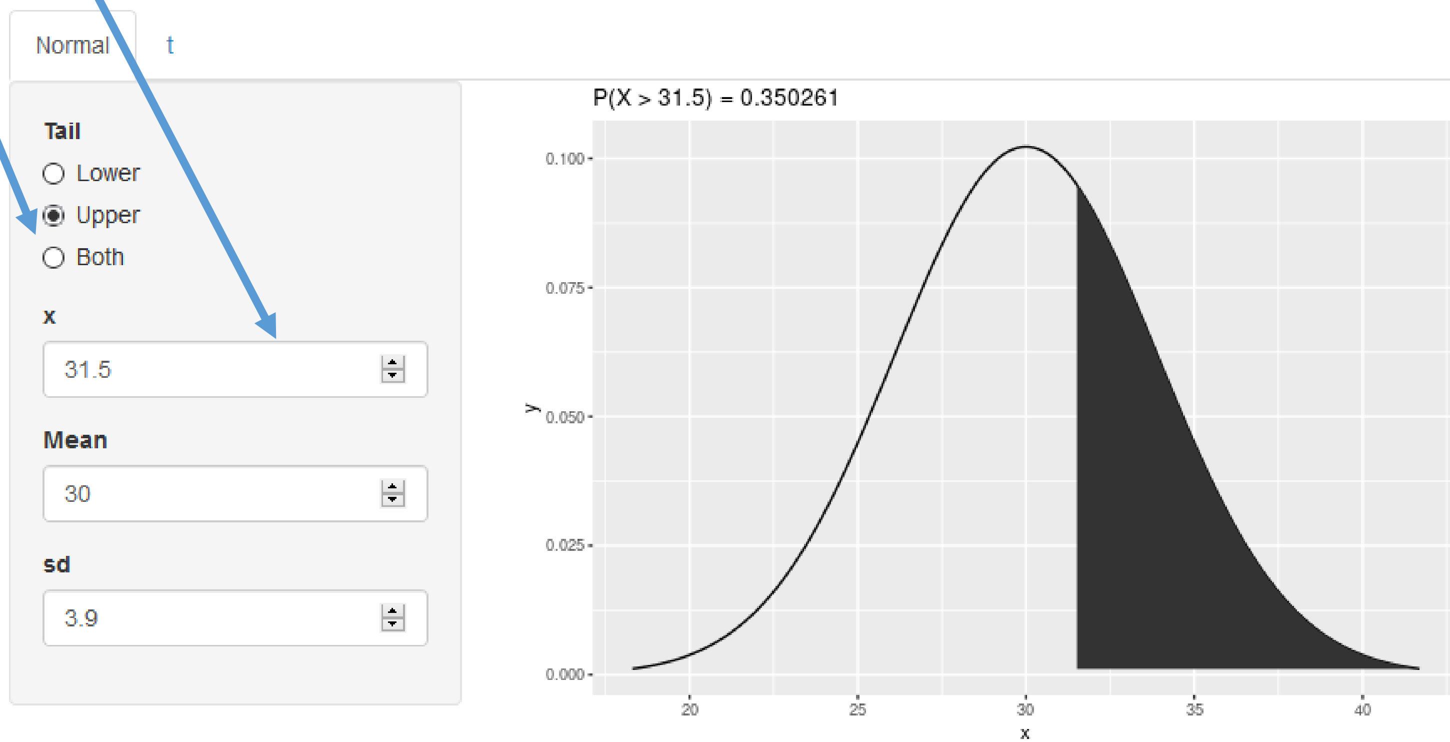
Sampling distribution – probability calculation

$x \equiv$ no. hours worked per week

$$P(x > 31.5) = 0.35026$$

Upper as probability is greater than x

Put the value of x in this box



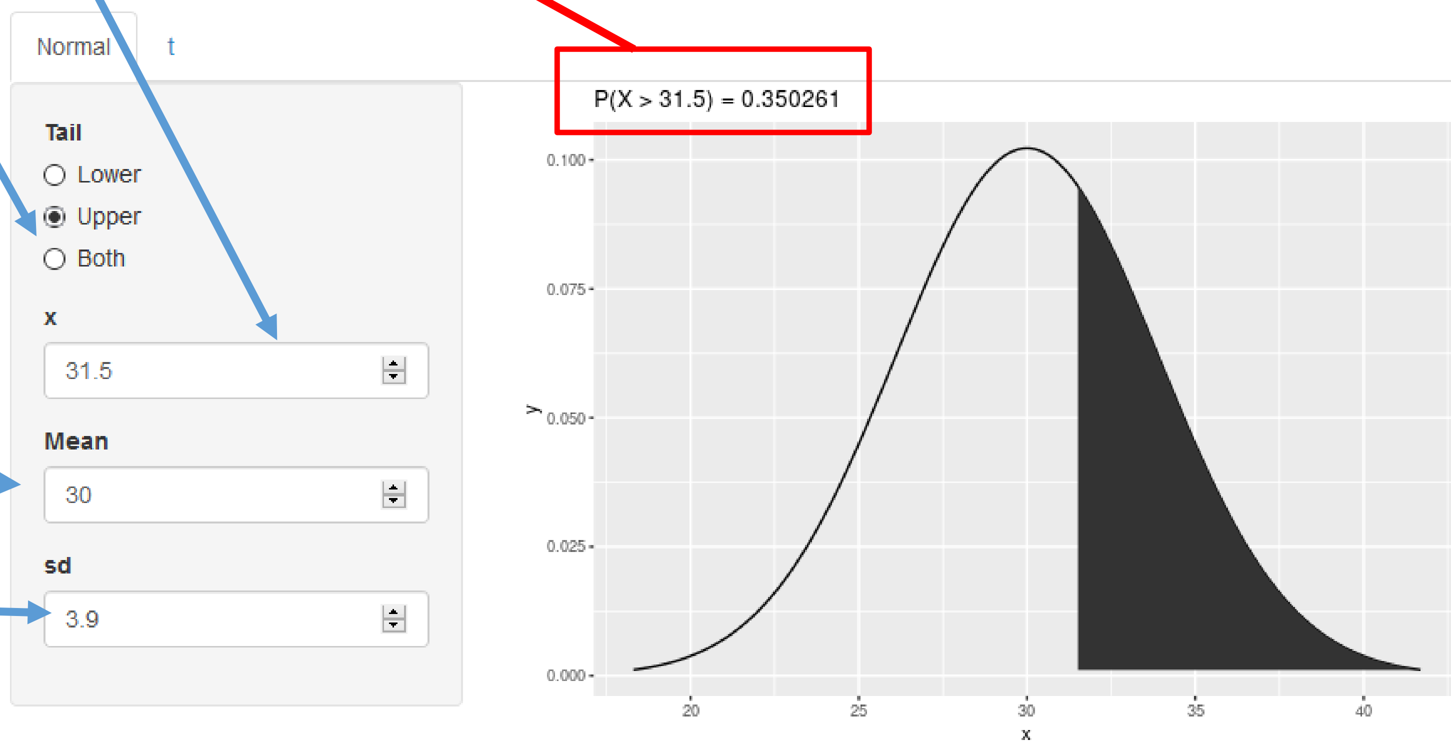
Sampling distribution – probability calculation

$x \equiv$ no. hours worked per week

$$P(x > 31.5) = 0.35026$$

Answer – always check with the graph that the area shaded is the one you think should be

Use the tables of the normal distribution with a mean of 30 and standard deviation of 3.9



Sampling distribution – probability calculation

- Example:
  - In a large company, the number of hours that full-time employees work in any week is known to be Normally distributed with $\mu = 30$h and standard deviation $\sigma = 3.9$h. What is the probability that the **mean** number of hours worked per week for 10 randomly selected employees is less than 31.5?

n

Mean number of hours, $\bar{x}$, is a random variable.

Its distribution is normal, with mean $\mu = 30$ and standard error:

$$\frac{\sigma}{\sqrt{N}} = \frac{3.9}{\sqrt{10}} = 1.2333$$

$$P(\bar{X} < 31.5)$$

Use the tables of the normal distribution with a mean of 30 and standard deviation of 1.2333

Sampling distribution – probability calculation

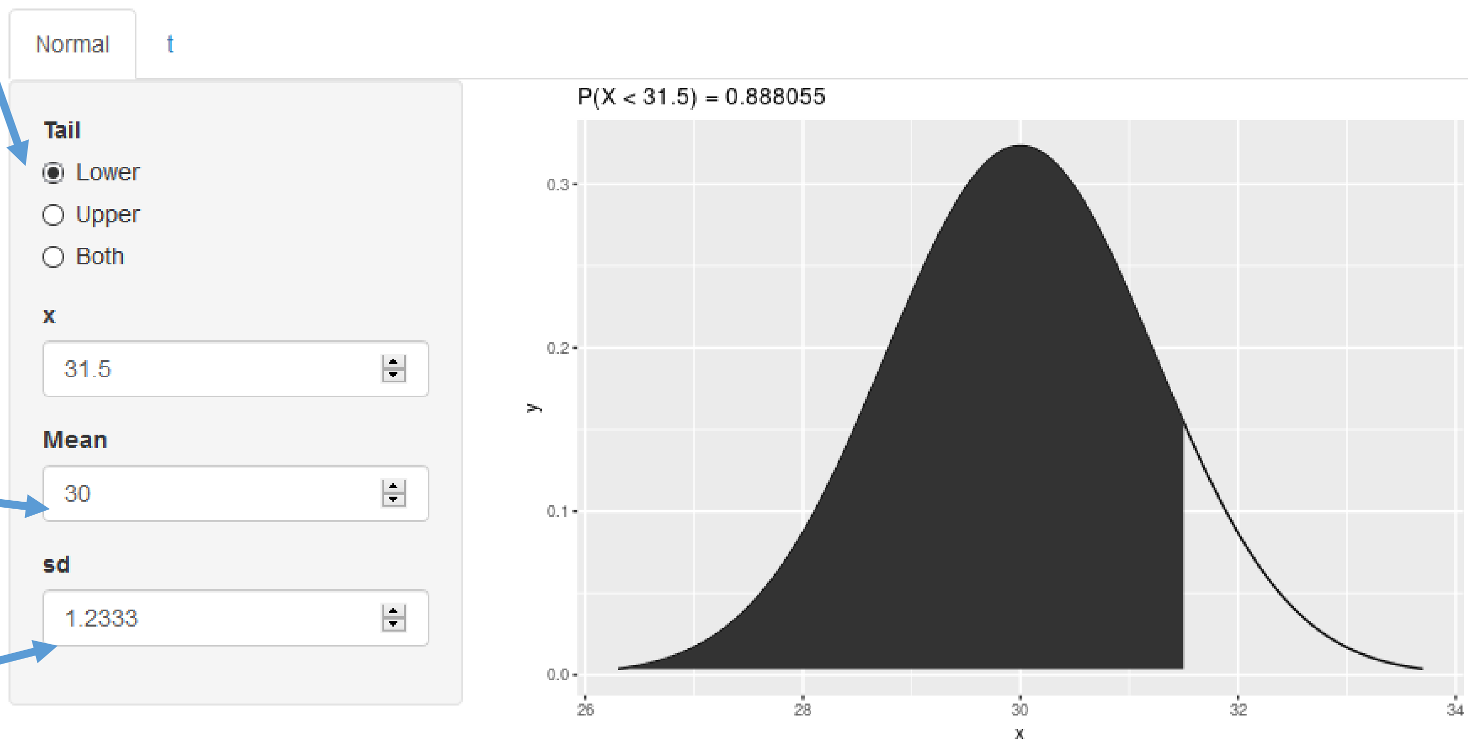
Mean number of hours, $\bar{X}$, is a random variable with a normal distribution.

mean $\mu = 30$ and standard error: $\frac{\sigma}{\sqrt{N}} = \frac{3.9}{\sqrt{10}} = 1.2333$

$$P(\bar{X} < 31.5) = 0.8881$$

Lower as probability is less than x

Use the tables of the normal distribution with a mean of 30 and standard deviation of 1.2333



Key Points

- The sampling distribution of the sample mean follows a normal distribution in large samples even if the original variable does not have a normal distribution
- The mean of the sampling distribution of the mean is the population mean

$$\mu_{\bar{X}} = \mu$$

- The standard deviation of the sampling distribution is known as the standard error of the sample mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- If the original variable is normally distributed the distribution of the sample mean is normal irrespective of the sample size