# ECO239

Week 13

## Practice

Body temperatures of healthy humans are distributed nearly normally with mean 98.2°F and standard deviation 0.73°F. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3°F

B. 99.1°F

C. 99.4°F

D. 99.6°F

# Practice

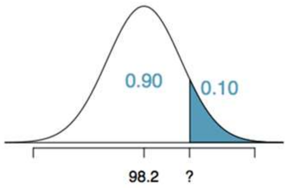
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

*B. 99.1ºF*

C. 99.4ºF

D. 99.6ºF




# Practice

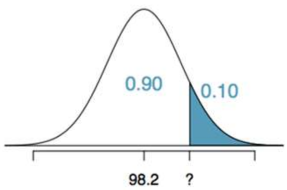
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

*B. 99.1ºF*

C. 99.4ºF

D. 99.6ºF




| Z | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|
| 1.0 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |

# Practice

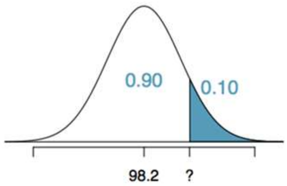
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

*B. 99.1ºF*

C. 99.4ºF

D. 99.6ºF




| Z | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|
| 1.0 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |

$$P(X > x) \quad = \quad 0.10 \rightarrow P(Z < 1.28) = 0.90$$

# Practice

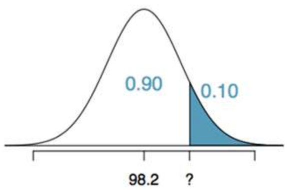
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

*B. 99.1ºF*

C. 99.4ºF

D. 99.6ºF




| Z | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|
| 1.0 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |

$$P(X > x) \quad = \quad 0.10 \rightarrow P(Z < 1.28) = 0.90$$

$$Z \quad = \quad \frac{obs \; - \; mean}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

# Practice

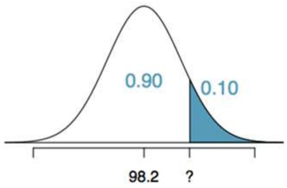
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

*B. 99.1ºF*

C. 99.4ºF

D. 99.6ºF




| Z | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|
| 1.0 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |

$$P(X > x) = 0.10 \rightarrow P(Z < 1.28) = 0.90$$

$$Z = \frac{obs - mean}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

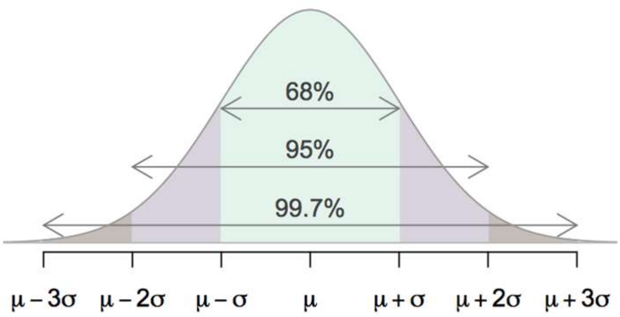
$$x = (1.28 \times 0.73) + 98.2 = 99.1$$

# 68-95-99.7 Rule

For nearly normally distributed data,

- about 68% falls within 1 SD of the mean,
- about 95% falls within 2 SD of the mean,
- about 99.7% falls within 3 SD of the mean.

It is possible for observations to fall 4, 5, or more standard deviations away from the mean, but these occurrences are very rare if the data are nearly normal.



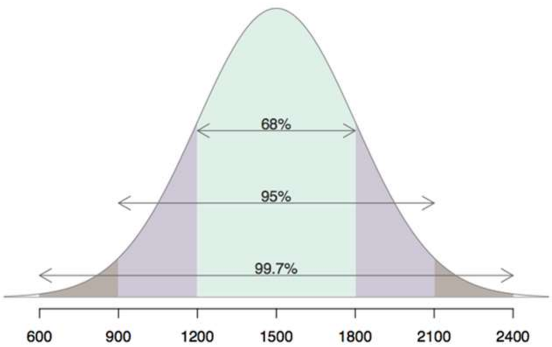
# Describing variability using the 68-95-99.7 Rule

SAT scores are distributed nearly normally with mean 1500 and standard deviation 300.

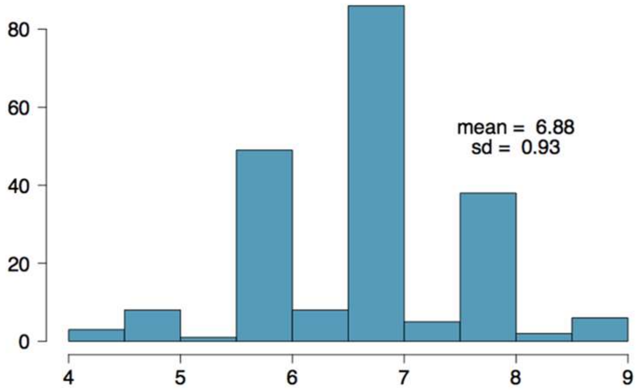
# Describing variability using the 68-95-99.7 Rule

SAT scores are distributed nearly normally with mean 1500 and standard deviation 300.

- ~68% of students score between 1200 and 1800 on the SAT.
- ~95% of students score between 900 and 2100 on the SAT.
- ~$99.7% of students score between 600 and 2400 on the SAT.

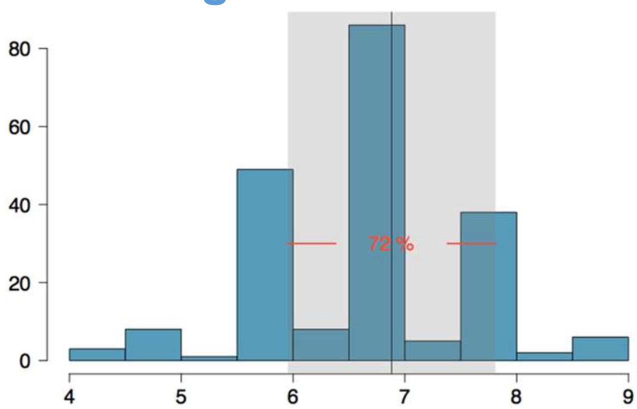


## Number of hours of sleep on school nights



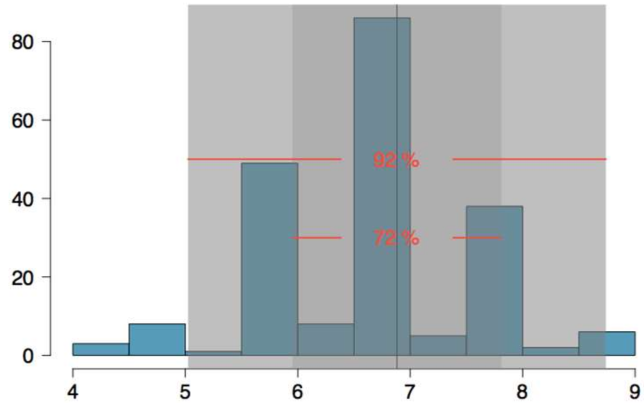

- Mean = 6.88 hours, SD = 0.92 hrs

## Number of hours of sleep on school nights



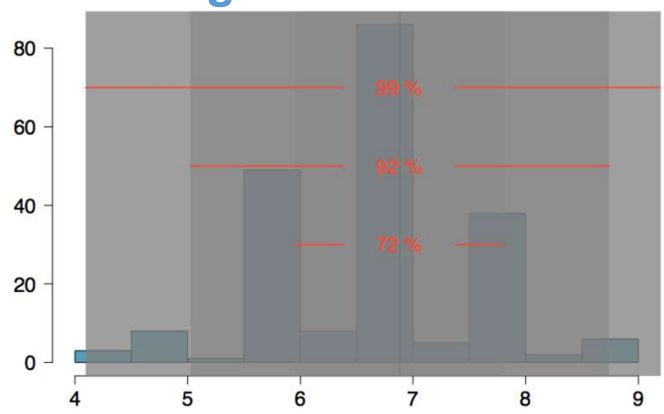

- Mean = 6.88 hours, SD = 0.92 hrs
- 72% of the data are within 1 SD of the mean: 6.88 ± 0.93

## Number of hours of sleep on school nights




- Mean = 6.88 hours, SD = 0.92 hrs
- 72% of the data are within 1 SD of the mean: 6.88 ± 0.93
- 92% of the data are within 1 SD of the mean: 6.88 ± 2 x 0.93

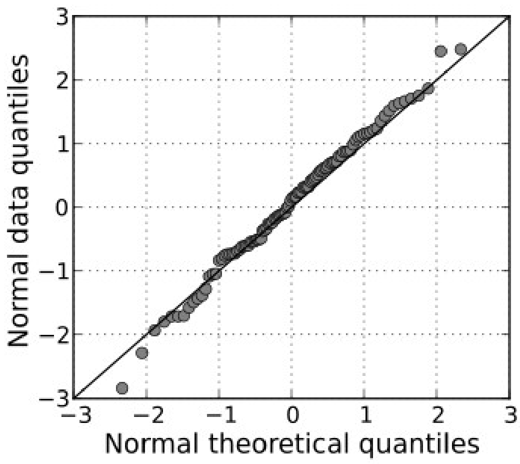
## Number of hours of sleep on school nights




- Mean = 6.88 hours, SD = 0.92 hrs
- 72% of the data are within 1 SD of the mean: 6.88 ± 0.93
- 92% of the data are within 2 SD of the mean: 6.88 ± 2 x 0.93
- 99% of the data are within 3 SD of the mean: 6.88 ± 3 x 0.93

# Evaluating the normal approximation

## Q-Q (Quartile-Quartile) plot




- The linearity of the points suggests that the data are normally distributed.

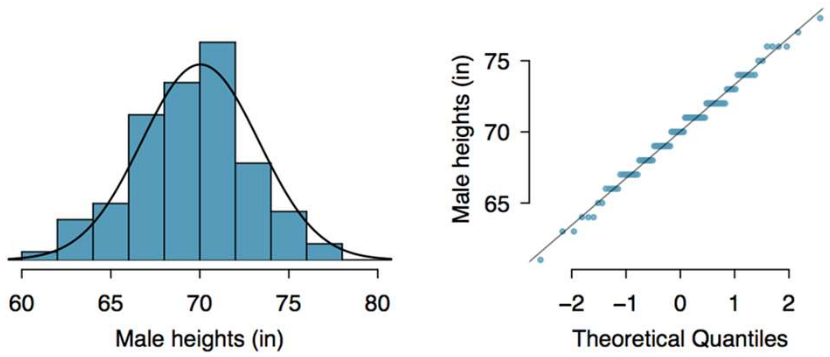
# Q-Q Plot (Quartile-Quartile Plot)

- In R, **qqnorm( )** function can be used to create a Quantile-Quantile plot evaluating the fit of sample data to the normal distribution.

- More generally, the **qqplot( )** function creates a Quantile-Quantile plot for any theoretical distribution to test if two data sets come from populations with a common distribution.

# Normal probability plot

A histogram and *normal probability plot* of a sample of 100 male heights.



## Anatomy of a normal probability plot

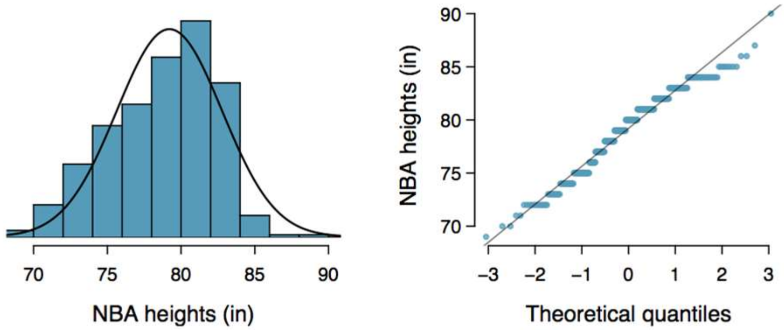
- Data are plotted on the y-axis of a normal probability plot, and theoretical quantiles (following a normal distribution) on the x-axis.
- If there is a linear relationship in the plot, then the data follow a nearly normal distribution.
- Constructing a normal probability plot requires calculating percentiles and corresponding z-scores for each observation, which is tedious. Therefore we generally rely on software when making these plots.

## How to draw a normal probability plot by hand

1. Arrange x-values in ascending order.
2. Calculate $fi = \frac{i-0.375}{n+0.25}$, where I is the position of the data value in the ordered list and n is the number of observations.
3. Find the z-score for each fi.
4. Plot x-values on the horizontal axis and the corresponding z-score on the vertical axis.

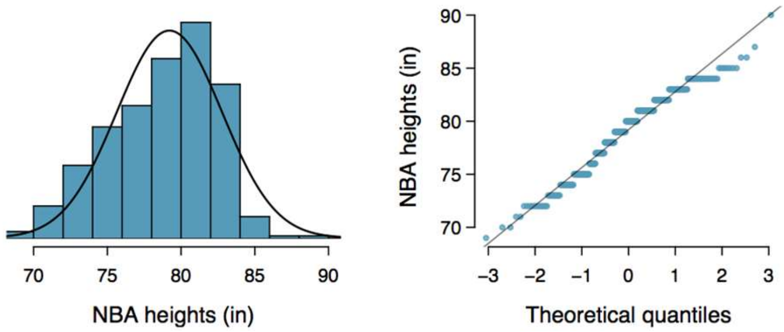
# Practice

Below is a histogram and normal probability plot for the NBA heights from the 2008-2009 season. Do these data appear to follow a normal distribution?



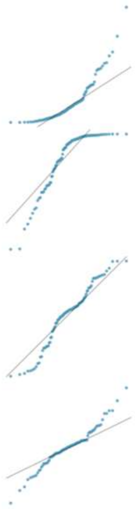

Practice

Below is a histogram and normal probability plot for the NBA heights from the 2008-2009 season. Do these data appear to follow a normal distribution?




Why do the points on the normal probability have jumps?

## Normal probability plot and skewness



Right skew - Points bend up and to the left of the line.

Left skew - Points bend down and to the right of the line.

Short tails (narrower than the normal distribution) - Points follow an S shaped-curve.

Long tails (wider than the normal distribution) - Points start below the line, bend to follow it, and end above it.

- Normal Probability Plot

(Watch this at home)

## Normal Distribution Approximation for Binomial Distribution

- Used to compute probabilities for large sample sizes when Binomial tables are not available.
- Binomial Distribution

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

- Good approximation if np(1-p)> 9

where $E(x) = np, Var(x) = np(1-p)$

$$Z = \frac{x - E(x)}{\sqrt{Var(x)}} = \frac{x - np}{\sqrt{np(1-p)}}$$

Review:
$np \leq 7$ , use Poisson Approximation for Binomial Distribution.

- X~ Bin(n,p). However, if np(1-p)>9, then treat as X~ N(np, sqrt[np(1-p)]).

$$Z = \frac{x - E(x)}{\sqrt{Var(x)}} = \frac{x - np}{\sqrt{np(1-p)}}$$

- Probabilities can be solved as

$$P(a \leq x \leq b) = p\left(\frac{a-np}{\sqrt{np(1-p)}} \leq \frac{x-np}{\sqrt{np(1-p)}} \leq \frac{b-np}{\sqrt{np(1-p)}}\right)$$

# Practice

- A salesman makes initial phone contact, then visit their homes if she assesses it is worthwhile. She knows that 40% of phone calls lead to follow-up visits.

Q: If she contact 100 people by phone, what is the probability that between 45 and 50 home visits (including 45 and 50) will result?

1. Direct method using Binomial Distribution.
2. Using Normal Approximation Method.

# Practice

- A salesman makes initial phone contact, then visit their homes if she assesses it is worthwhile. She knows that 40% of phone calls lead to follow-up visits.

Q: If she contact 100 people by phone, what is the probability that between 45 and 50 home visits will result?

**1: Direct method (use Binomial Distribution)**

$$P(45 \le x \le 50) = P(x = 45) + P(x = 46) + \cdots + P(x = 50)$$
$$= \frac{100!}{45!\,55!} 0.4^{45} 0.6^{55} + \frac{100!}{46!\,54!} 0.4^{46} 0.6^{54} \ldots$$
$$+ \frac{100!}{50!\,50!} 0.4^{50} 0.6^{50} = ???$$

Or by using Binomial Table:
F(50|n=100,p=0.4)-F(45|n=100,p=0.4) = ?

- A salesman makes initial phone contact, then visit their homes if she assesses it is worthwhile. She knows that 40% of phone calls lead to follow-up visits.

Q: If she contact 100 people by phone, what is the probability that between 45 and 50 home visits will result?

**2: Using Normal Approximation** [ np(1-p)=100*0.4*0.6=24 > 9 ]

E(x)=np=100*0.4 = 40, Var(x)=np(1-p)=100*0.4*0.6=24.

$$P(45 \leq x \leq 50) = P\left(\frac{45-40}{\sqrt{24}} \leq z \leq \frac{50-40}{\sqrt{24}}\right)$$
$$= P(1.02 \leq z \leq 2.04) = F(2.04) - F(1.02) =$$

# Exponential Distribution

- Used for modeling waiting time or queuing problems.
- Positive Random Variable: t

$\Rightarrow$t: number of time units until next occurrence.

- Depends on a single parameter, $\lambda > 0$

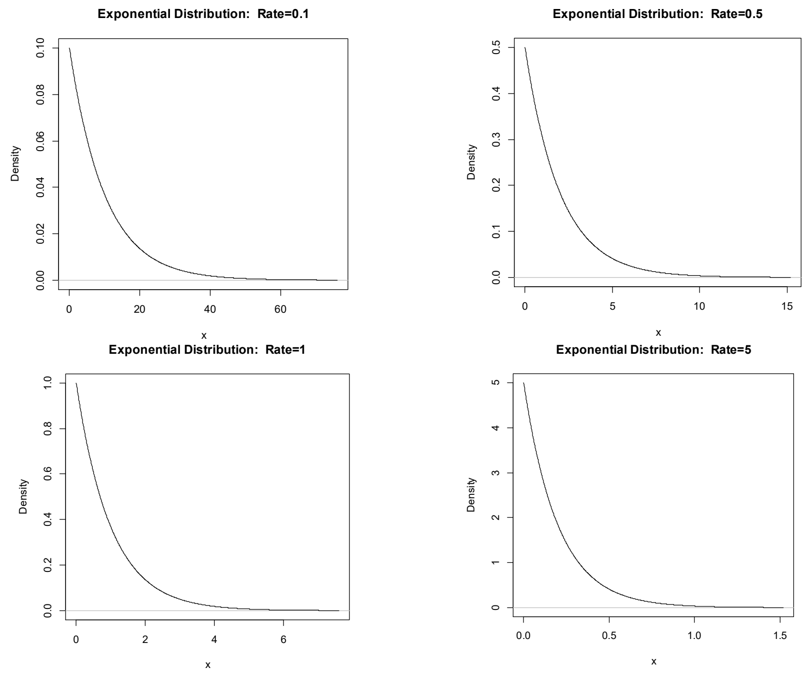
$\Rightarrow\lambda$: mean number of occurrences per time unit.

- Distribution is not symmetric

PDF: $$f(t) = \lambda e^{-\lambda t} \text{ for } t \geq 0$$

CDF: $$F(t) = 1 - e^{-\lambda t} \text{ for } t \geq 0$$

* Use CDF for the calculation of Probabilities!




## Comparing with Poisson Distribution

- Poisson: probability of **x successes** during a time unit (e.g. 5 minutes): RV (X) is # of successes (discrete variable).

$$P(X) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

- Exponential: probability that **a success** will occur during **an interval of time t**. RV (t) is time (continuous variable).

$$f(t) = \lambda e^{-\lambda t} \text{ for } t \geq 0$$

## Practice

$$F(t) = 1 - e^{-\lambda t} \text{ for } t \geq 0$$

- Service time at a repair shop can be modeled by an exponential distribution with mean service time of 5 minutes. What is the probability that a customer service time will take longer than 10 minutes?

- t: service time in minute
- λ: 1/5 service per minute (<= 1 service per 5 min.)

## Practice

$$F(t) = 1 - e^{-\lambda t} \text{ for } t \geq 0$$

- A professor sees students during his office hour. Time spent with students follow an exponential distribution with mean 10 minutes.

Q1: Find the probability that a given student spends less than 20 mins with the professor.

Q2: Find the probability that a given student spends more than 5 mins with the professor.

- A professor sees students during his office hour. Time spent with students follow an exponential distribution with mean 10 minutes.

Q1: Find the probability that a given student spends less than 20 mins with the professor.

Q2: Find the probability that a given student spends more than 5 mins with the professor.

$\lambda$ = 1/10 meeting per minute.
(<=1 meeting per 10 mins)

Q1: P(t<20)=1-e^(-0.1*20)=1-e^-2=0.865

Q2: P(t>5)=1-[1-e^(-0.1*5)]=e^-0.5=0.607

## Jointly Distributed Continuous R.V.

- Joint Cumulative Distribution Function

Let X1 and X2 be continuous R.V.

1. Joint CDF: $F(X1, X2) = P(X1<x1 \cap X2<x2)$
2. Marginal distribution function : $F(X1)=P(X1<x1)$
3. RVs are independent iff $F(X1,X2)=F(x1)F(x2)$
4. $Cov(X,Y)=E[(X-\mu x)(Y-\mu y)] = E(XY)-\mu x \mu y$.
5. $$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

$$W = aX + bY$$

$$E(W) = a\mu_X + b\mu_Y$$

$$\begin{aligned} Var(W) &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X, Y) \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 \\ &\quad + 2abCor(X, Y)\sigma_X\sigma_Y \end{aligned}$$

## Practice

Px~N(25, 9), Py~N(40,11)
- Nx=20 (# of Stock X)
- NY=30 (# of Stock Y)
- Cor(Px,Py)=-0.40.

$$\begin{aligned} Var(W) &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X, Y) \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 \\ &\quad + 2abCor(X, Y)\sigma_X\sigma_Y \end{aligned}$$

Q: find the probability that the portfolio (W=20Px+30Py) value exceeds 2000.

# Practice

Px~N(25, 9), Py~N(40,11)

- Nx=20 (# of Stock X)
- NY=30 (# of Stock Y)
- Cor(X,Y)=-0.40.

Q: find the probability that the portfolio (W=20Px+30Py) value exceeds 2000.

E(W)=20*25+30*40=1700
Var(W)=400*81+900*121+2*20*30*(-0.4)*9*11 = 98780
Stdev(W)=306.24

P(W>2000)=P(z> (2000-1700)/306.24) = P(z>0.98)=1-F(0.98)=1-0.8365=0.1635.
The probability for the portfolio value to exceed 2000 is 16.35%.