

Descriptive statistics:
Sample mean and
variance

Linear Functions of Random Variables

- A **function of random variables** is itself a **random variable**.
- A function of random variables can be formed by either **linear** or **nonlinear relationships**. We start with linear functions.
- Given random variables $X_1, X_2, \dots, X_p$ and constants $c_1, c_2, \dots, c_p$
$$Y = c_1X_1 + c_2X_2 + \dots + c_pX_p \quad (5-24)$$
is a linear combination of $X_1, X_2, \dots, X_p$.

Mean & Variance of a Linear Function

$$Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$$

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \dots + c_pE(X_p) \quad (5-25)$$

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \dots + c_p^2V(X_p) + 2 \sum_{i < j} \sum c_i c_j \text{cov}(X_i, X_j) \quad (5-26)$$

If $X_1, X_2, \dots, X_p$ are **independent**, then $\text{cov}(X_i, X_j) = 0$,

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \dots + c_p^2V(X_p) \quad (5-27)$$

Example 5-31: Error Propagation

A semiconductor product consists of three layers. The variances of the thickness of each layer is 25, 40 and 30 nm². What is the variance of the finished product?

Example 5-31: Error Propagation

A semiconductor product consists of three layers. The variances of the thickness of each layer is 25, 40 and 30 nm². What is the variance of the finished product?

Answer:

IMPORTANT

$$X = X_1 + X_2 + X_3$$

Add variations
not
standard
deviations

$$V(X) = \sum_{i=1}^3 V(X_i) = 25 + 40 + 30 = 95 \text{ nm}^2$$

$$SD(X) = \sqrt{95} = 9.7 \text{ nm}$$

If adding SDs one would get $\sqrt{25} \text{ nm} + \sqrt{40} \text{ nm} + \sqrt{30} \text{ nm} = 16.18 \text{ nm}$

Some Definitions

- The random variables $X_1, X_2, \dots, X_n$ are a **random sample** of **size n** if:
 - a) The X_i are **independent** random variables.
 - b) Every X_i has **the same probability distribution**.
Such $X_1, X_2, \dots, X_n$ are also called independent and identically distributed (or **i. i. d.**) random variables
- A **statistic** is any function of the observations in a random sample.
- The probability distribution of a statistic is called a **sampling distribution**.

Statistic #1: Sample Mean

If the n observations in a random sample are denoted by $x_1, x_2, \dots, x_n$, the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (6-1)$$

IMPORTANT:

Sample mean $\bar{x}$ is drawn from a random variable

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$E(\bar{X}) = \frac{n \cdot E(X_i)}{n} = \frac{n \cdot \mu}{n} = \mu$$

$$V(\bar{X}) = \frac{n \cdot V(X_i)}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\text{Stand. dev.}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

If $X_1, X_2, \dots, X_n$ is a random sample of size n is taken from a population (either finite or infinite) with mean μ and finite variance σ^2 , and if $\bar{X}$ is the sample mean, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (7-1)$$

for large n , is the standard normal distribution.

If $X_1, X_2, \dots, X_n$ are themselves normally distributed
- for any n

Sampling Distributions of Sample Means

Figure 7-1 Distributions of average scores from throwing dice.

Mean = $(6+1)/2=3.5$

$\text{Sigma}^2 = [(6-1+1)^2-1]/12=2.92$

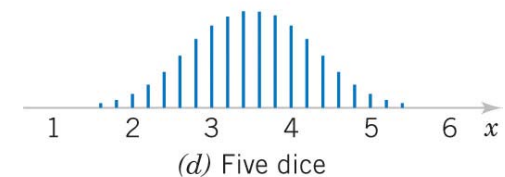
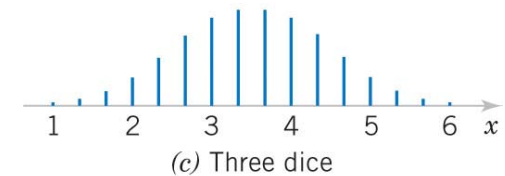
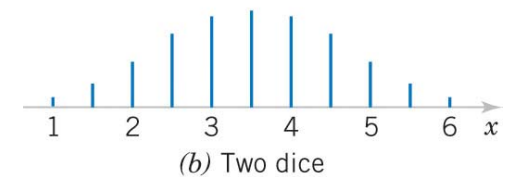
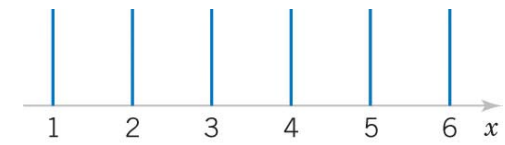
$\text{Sigma}=1.71$

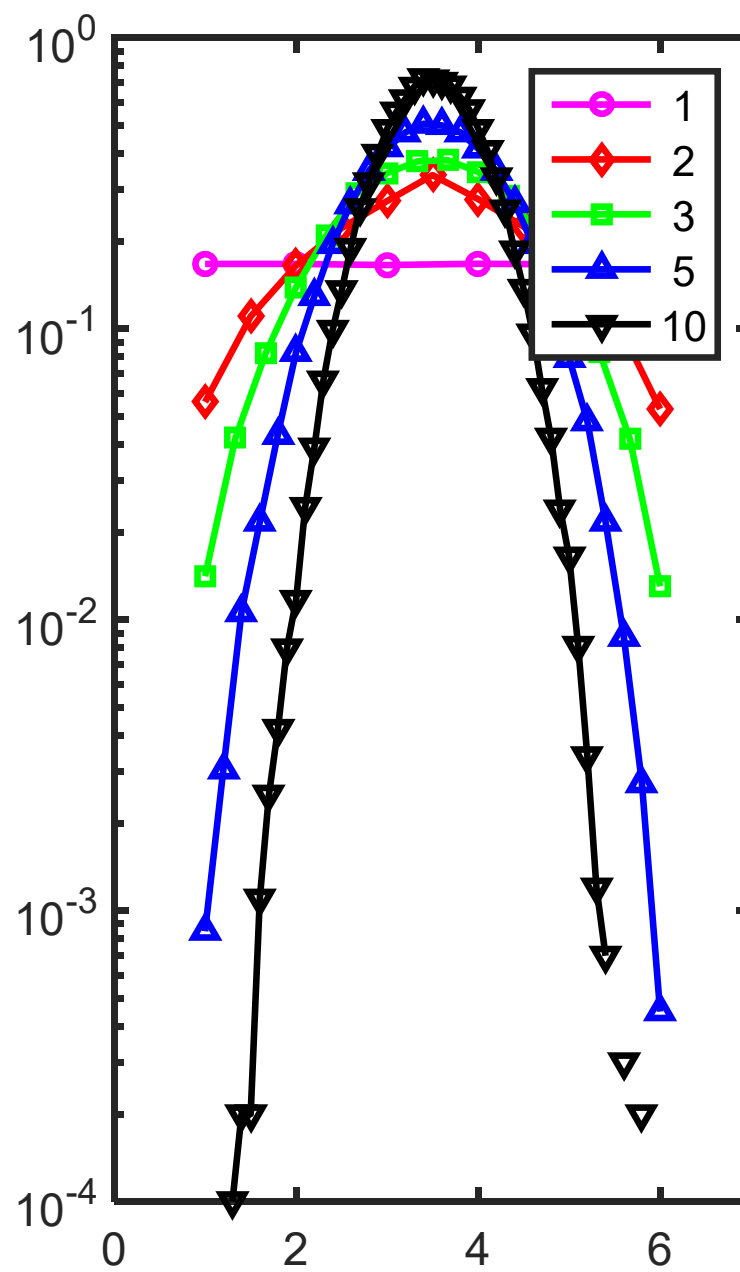
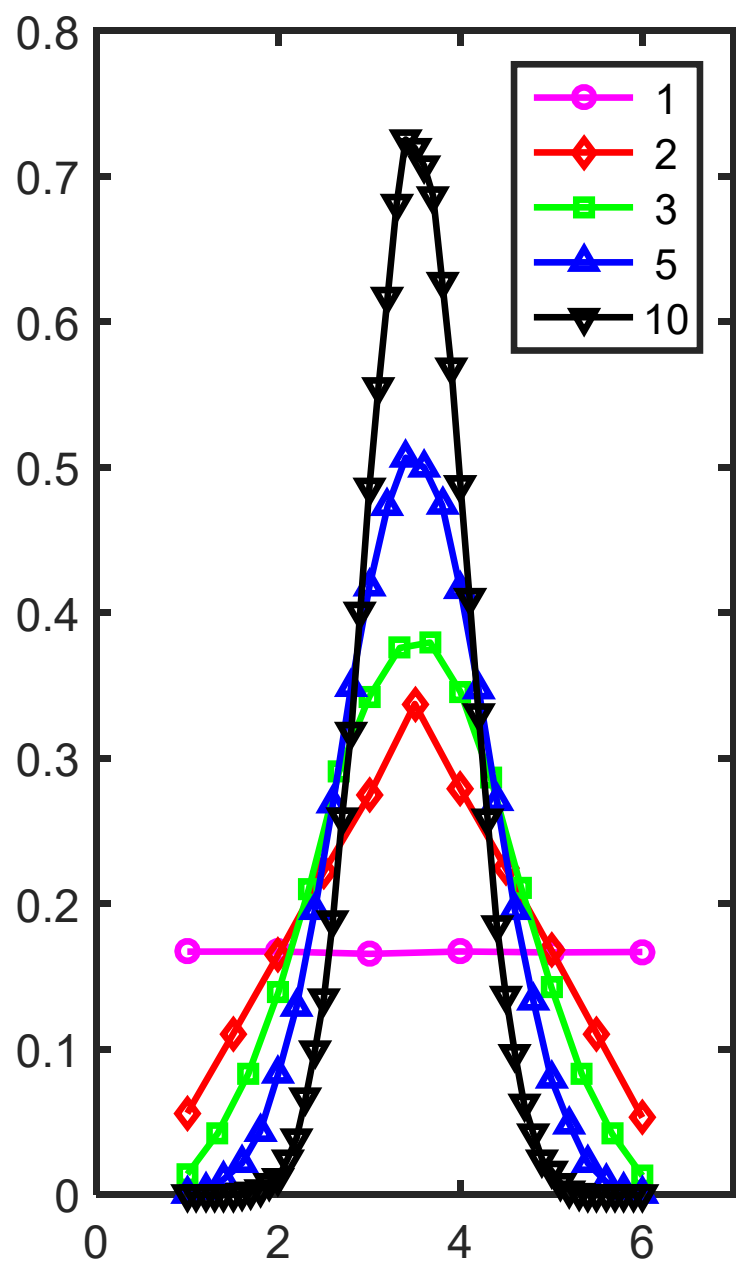
Formulas

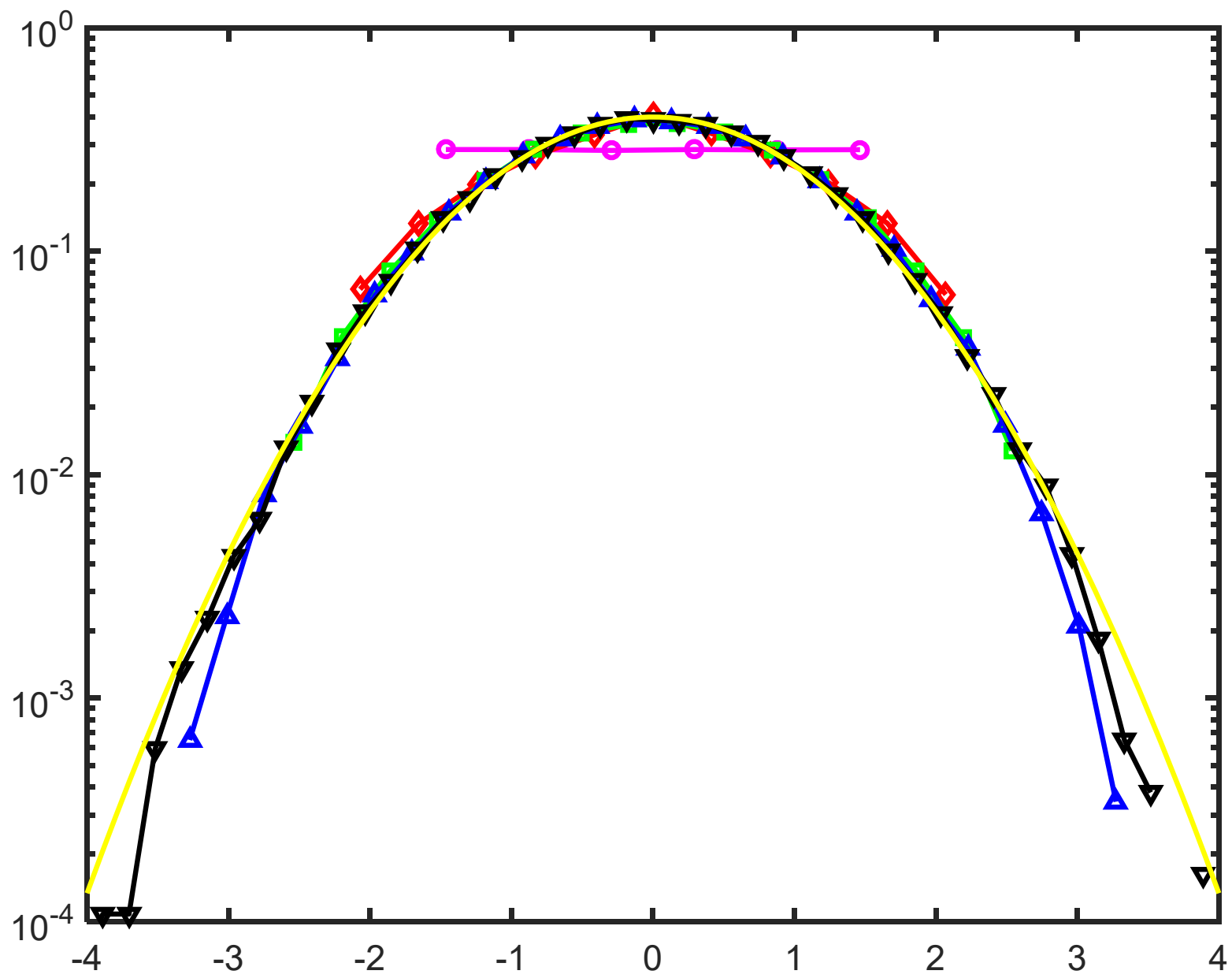
$$\mu = \frac{b+a}{2}$$

$$\sigma_X^2 = \frac{(b-a+1)^2 - 1}{12}$$

$$\sigma_{\bar{X}}^2 = \sigma_X^2 / n$$







Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. What is the approximate probability that a random sample of $n = 25$ resistors will have an average resistance of less than 95 ohms?

Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. What is the approximate probability that a random sample of $n = 25$ resistors will have an average resistance of less than 95 ohms?

$$\mu = 100 \text{ ohms}, \quad \sigma = 10 \text{ ohms}, \quad n = 25$$

$$\mu_{\bar{x}} = \mu; \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2 \text{ ohms}$$

$$Z_{\bar{x}} = \frac{95 - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{95 - 100}{2} = -2.5$$

$$\text{Prob}(\bar{X} < 95) = \Phi(Z_{\bar{x}}) = \Phi(-2.5) =$$

$$1 - \Phi(2.5) = 1 - 0.9938 = 0.0062$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. What is the approximate probability that a random sample of $n = 25$ resistors will have an average resistance of less than 95 ohms?

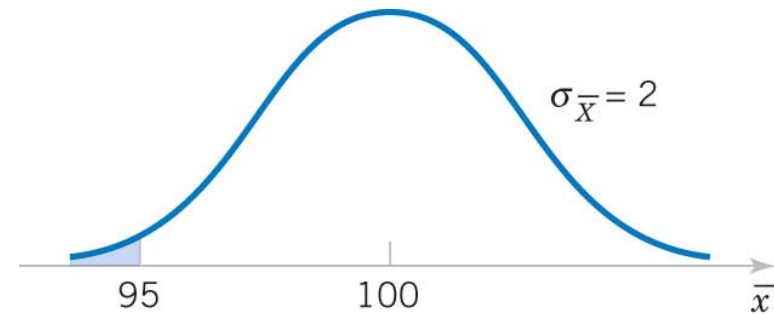


Figure 7-2 Desired probability is shaded

Answer:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2.0$$

$$\begin{aligned}\Phi\left(\frac{\bar{X} - \mu}{\sigma_{\bar{X}}}\right) &= \Phi\left(\frac{95 - 100}{2}\right) \\ &= \Phi(-2.5) = 1 - \Phi(2.5) \\ &= 1 - 0.9938 = 0.0062\end{aligned}$$

Two Populations

We have two independent populations. Two samples of sizes n_1 and n_2 respectively were drawn from these populations. What is the distribution of the difference of their sample means?

The sampling distribution of $\bar{X}_1 - \bar{X}_2$ has the following mean and variance:

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{X}_1 - \bar{X}_2}^2 = \sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Sampling Distribution of a Difference in Sample Means

- If we have two independent populations with means μ_1 and μ_2 , and variances σ_1^2 and σ_2^2 ,
- And if $\bar{X}_1$ and $\bar{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations:
- Then the sampling distribution of:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (7-4)$$

is approximately standard normal, if the conditions of the central limit theorem apply.

- If the two populations are normal, then the sampling distribution is exactly standard normal.

Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a random variable with $\mu_{\text{old}}=5000$ hours and $\sigma_{\text{old}}=40$ hours (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters to $\mu_{\text{new}}=5050$ hours and $\sigma_{\text{new}}=30$ hours (new).

Random samples of 16 components manufactured using “old” process and 25 components using “new” process are chosen.

What is the probability new sample mean is at least 25 hours longer than old?

Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a random variable with $\mu_{old}=5000$ hours and $\sigma_{old}=40$ hours (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters to $\mu_{new}=5050$ hours and $\sigma_{new}=30$ hours (new).

Random samples of 16 components manufactured using "old" process and 25 components using "new" process are chosen.

What is the probability new sample mean is at least 25 hours longer than old?

$$\sigma_{\bar{x}_{old}} = \frac{\sigma_{old}}{\sqrt{16}} = 10 \text{ hrs}$$

$$\sigma_{\bar{x}_{new}} = \frac{\sigma_{new}}{\sqrt{25}} = 6 \text{ hrs}$$

$$\sigma_{TOT} = \sqrt{\sigma_{\bar{x}_{old}}^2 + \sigma_{\bar{x}_{new}}^2} =$$

$$= \sqrt{100 + 36} \approx 11.7 \text{ hrs}$$

$$\mu_{new} - \mu_{old} = 50 \text{ hrs}$$

$$z = \frac{25 - (50)}{11.7} = -2.14$$
$$\text{Prob}(z > -2.14) = 0.9840$$

Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a normal-distributed random variable with parameters shown (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters μ and σ as shown (new).

Random samples are selected from the “old” process and “new” process as shown.

What is the probability new sample mean is at least 25 hours longer than old?

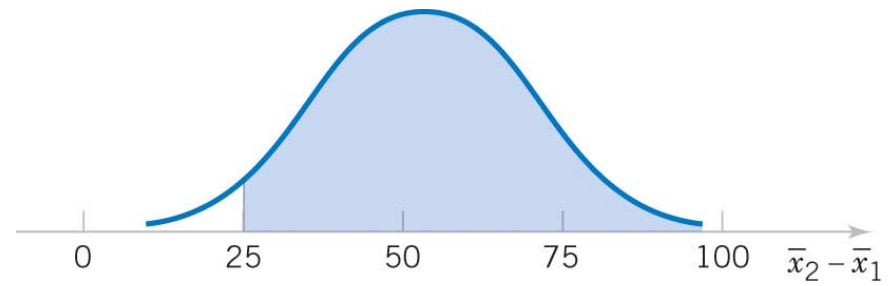


Figure 7-4 Sampling distribution of the sample mean difference.

	Process		
	Old (1)	New (2)	Diff (2-1)
$\mu =$	5,000	5,050	50
$\sigma =$	40	30	50
$n =$	16	25	
Calculations			
$s / \sqrt{n} =$	10	6	11.7
		$z =$	-2.14
$P(\bar{x}_2 - \bar{x}_1 > 25) = P(Z > z) =$			0.9840

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

Credit: XKCD
comics

QUESTIONS FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS IN MAY DO I FEEL DIZZY

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES

WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARIOUS PRIETIES
WHY ARE OLD KLINGONS DIFFERENT



WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN



WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS

WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY

WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD



WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

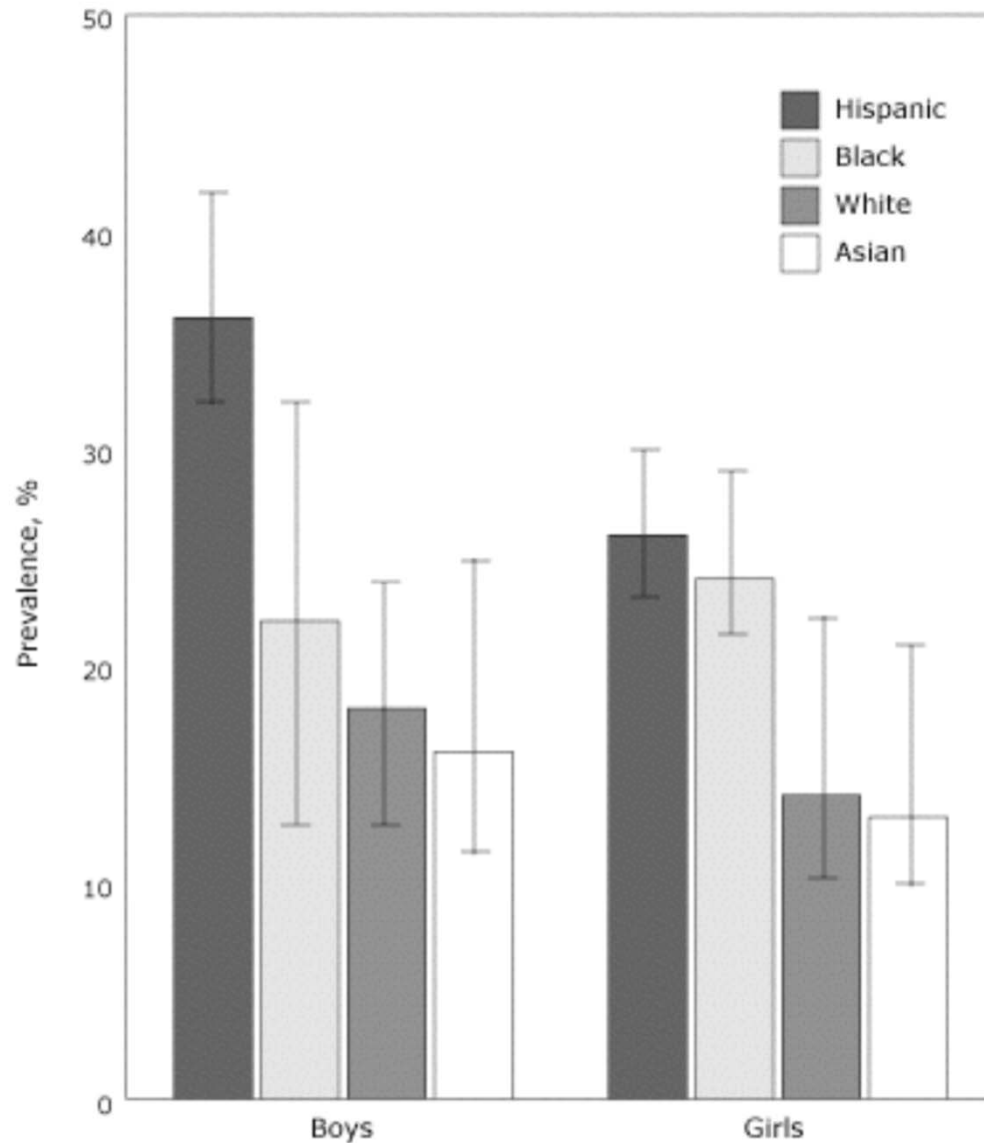
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE



WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND

Confidence intervals



Prevalence (with 95% CI bars) of obesity among New York City public elementary schoolchildren, by sex and race/ethnicity, 2003.

(source: CDC.GOV)

What do those bars actually mean?

ARTICLES

Patterns of somatic mutation in human cancer genomes

What does confidence interval mean?

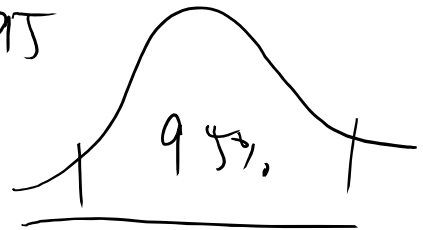
The numbers of passenger and driver mutations present can be estimated from these results (see Supplementary Methods). Of the 921 base substitutions in the primary screen, 763 (95% confidence interval, 675–858) are estimated to be passenger mutations. Therefore, the large majority of mutations found through sequencing cancer genomes are not implicated in cancer development, even when the search has been targeted to the coding regions of a gene family of high candidature. However, there are an estimated 158 driver mutations (95% confidence interval, 63–246), accounting for the observed positive selection pressure. These are estimated to be distributed in 119 genes (95% confidence interval, 52–149). The number of samples containing a driver mutation is estimated to be 66 (95% confidence interval, 36–77). The results, therefore, provide statistical evidence for a large set of mutated protein kinase genes implicated in the development of about one-third of the cancers studied.

Two-sided confidence intervals

- Calculated based on the sample $X_1, X_2, \dots, X_n$
- Characterized by:
 - lower- and upper- confidence limits L and R
 - the confidence coefficient $1-\alpha$
- Objective: for two-sided confidence interval, find L and R such that
 - $\text{Prob}(\mu > R) = \alpha/2$
 - $\text{Prob}(\mu < L) = \alpha/2$
 - Therefore, $\text{Prob}(L < \mu < R) = 1-\alpha$
- For one-sided confidence interval, say, lower bound of μ , find R that
 - $\text{Prob}(\mu > R) = \alpha$
- **Assume standard deviation sigma is known**

Consider $1 - \alpha = 95\% = 0.95$

$$\alpha = 0.05; \quad \frac{\alpha}{2} = 0.025$$



$$z_{\alpha/2} = 1.96 \rightarrow \text{Prob}(Z > z_{\alpha/2}) = \frac{\alpha}{2} = 2.5\% \quad 2.5\%$$

$$\text{Prob}\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\text{Prob}\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

For one sided lower bound on μ

$$\text{Prob}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \underline{\underline{z_{\alpha}}}\right) \rightarrow$$

$$\mu > \bar{X} - z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

$$z_{\alpha} = 1.65 <$$

$$z_{\alpha/2} = 1.96$$

Exercise

Ishikawa et al. (Journal of Bioscience and Bioengineering 2012) studied the force with which bacterial biofilms adhere to a solid surface.

Five measurements for a bacterial strain of *Acinetobacter* gave readings 2.69, 5.76, 2.67, 1.62, and 4.12 dyne-cm².

Assume that the standard deviation is known to be 0.66 dyne-cm²

- (a) Find 95% confidence interval for the mean adhesion force
- (b) If scientists want the width of the confidence interval to be below 0.55 dyne-cm² what the minimal number of samples should be?

Ishikawa et al. (Journal of Bioscience and Bioengineering 2012) studied the force with which bacterial biofilms adhere to a solid surface. Five measurements for a bacterial strain of Acinetobacter gave readings 2.69, 5.76, 2.67, 1.62, and 4.12 dyne-cm². Assume that the **standard deviation is known to be 0.66 dyne-cm²**

- (a) Find 95% confidence interval for the mean adhesion force
- (b) If scientists want the width of the confidence interval to be below 0.55 dyne-cm² what the minimal number of samples should be?

a) 95% CI for μ , $n = 5$ $\sigma = 0.66$ $\bar{x} = 3.372$, $z = 1.96$

$$\bar{x} - z\sigma / \sqrt{n} \leq \mu \leq \bar{x} + z\sigma / \sqrt{n}$$

$$3.372 - 1.96(0.66 / \sqrt{5}) \leq \mu \leq 3.372 + 1.96(0.66 / \sqrt{5})$$

$$2.79 \leq \mu \leq 3.95$$

b) Width is $2z\sigma / \sqrt{n} = 0.55$, therefore $n = [2z\sigma / 0.55]^2 = [2(1.96)(0.66) / 0.55]^2 = 22.13$
Round up to $n = 23$.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

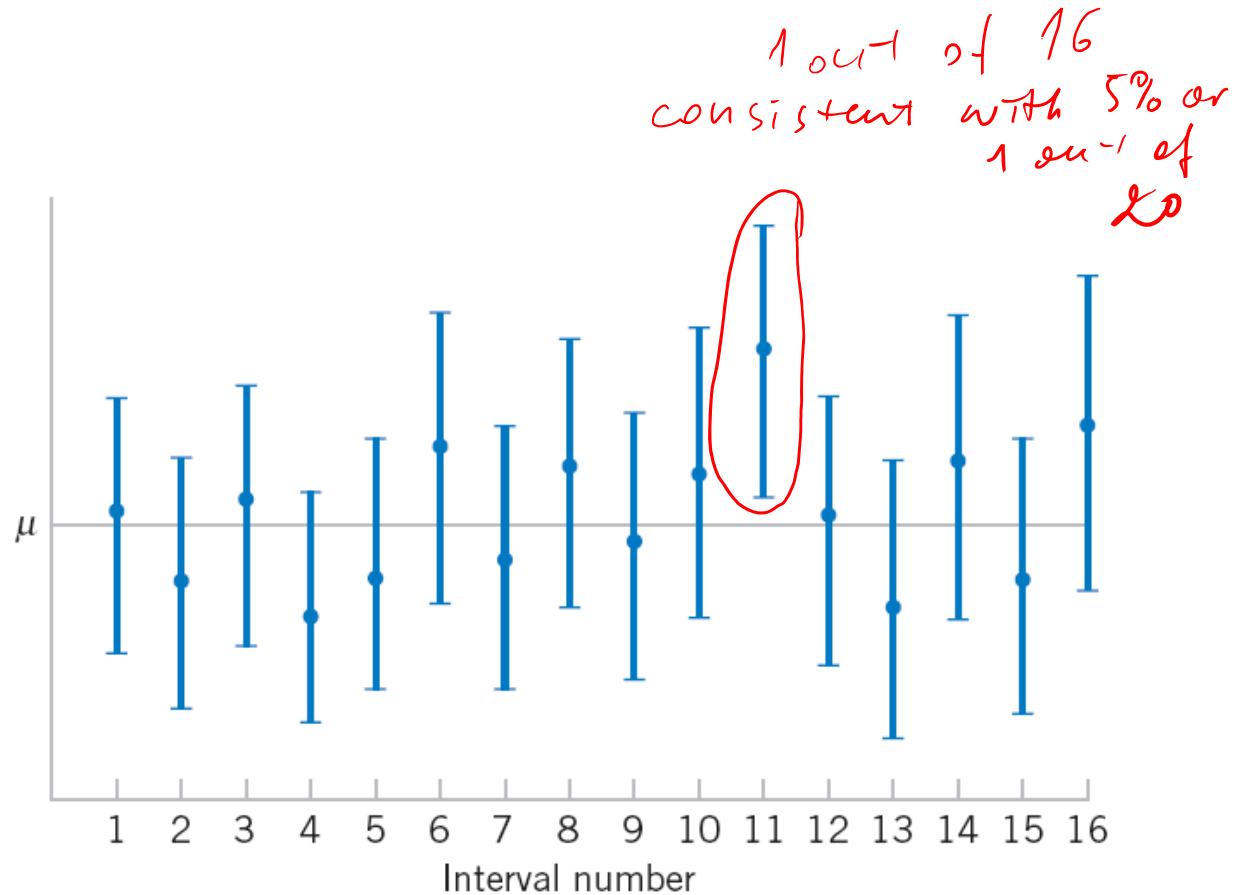


Figure 8-1 Repeated construction of a confidence interval for μ .

Figure 8-1 Repeated construction of a confidence interval for μ .

Matlab exercise

- 1000 labs measured average P53 gene expression.
- Each lab used a sample with $n=20$ drawn from the Gaussian distribution with $\mu=3$; $\sigma=2$;
- Each lab found 95% confidence estimates of the population mean μ **based on its own sample only**
- Count the number of labs, where the true population mean ($\mu=3$) lies **outside their 95% confidence bounds**
- You should get ~ 50 labs out of 1000 labs

How I did it

- `n=20; k_labs=1000;`
- `rand_table=2.*randn(n,k_labs)+3;`
- `sample_mean=mean(rand_table,1);`
- `CI_low=sample_mean-1.96.*2./sqrt(n);`
- `CI_high=sample_mean+1.96.*2./sqrt(n);`
- `k_above=sum(3>CI_high)`
- `k_below=sum(3<CI_low)`
- `figure; ndisp=100; errorbar(1:ndisp,
sample_mean(1:ndisp),
ones(ndisp,1).*1.96.*2./sqrt(n),'ko');`
- `hold on; plot(1:ndisp, 3.*ones(ndisp,1),'r-');`

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

Credit: XKCD
comics

QUESTIONS FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS IN MAY DO I FEEL DIZZY

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES

WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARIOUS PRIETIES
WHY ARE OLD KLINGONS DIFFERENT



WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN



WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS

WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY

WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD



WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE



WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND