Math 11
Calculus-Based Introductory Probability and Statistics

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

Today: Review session

- Review some of the main ideas from the course
- Practice problems in a timed setting

Warm Up

Suppose 33 fish are sampled from a lake. The mean of their lengths is 4.4 inches, and the standard deviation is 1.3 inches. You may assume that the 33 fish represent a random sample of all fish in the lake. Find a 90% confidence interval for the average length of all fish in the lake.

Warm Up

Suppose 33 fish are sampled from a lake. The mean of their lengths is 4.4 inches, and the standard deviation is 1.3 inches. You may assume that the 33 fish represent a random sample of all fish in the lake. Find a 90% confidence interval for the average length of all fish in the lake.

This problem is about a one-sample t-test.

Warm Up

Suppose 33 fish are sampled from a lake. The mean of their lengths is 4.4 inches, and the standard deviation is 1.3 inches. You may assume that the 33 fish represent a random sample of all fish in the lake. Find a 90% confidence interval for the average length of all fish in the lake.

This problem is about a one-sample t-test.

The C.I. is $\bar{x} \pm t^*_{n-1} \cdot SE_{\bar{x}}$.

Warm Up

Suppose 33 fish are sampled from a lake. The mean of their lengths is 4.4 inches, and the standard deviation is 1.3 inches. You may assume that the 33 fish represent a random sample of all fish in the lake. Find a 90% confidence interval for the average length of all fish in the lake.

one tail	0.100	0.050	0.025	0.010	0.005
two tails	0.200	0.100	0.050	0.020	0.010
df 31	1.31	1.70	2.04	2.45	2.74
32	1.31	1.69	2.04	2.45	2.74
33	1.31	1.69	2.03	2.44	2.73
34	1.31	1.69	2.03	2.44	2.73
35	1.31	1.69	2.03	2.44	2.72
36	1.31	1.69	2.03	2.43	2.72
37	1.30	1.69	2.03	2.43	2.72
38	1.30	1.69	2.02	2.43	2.71
39	1.30	1.68	2.02	2.43	2.71
40	1.30	1.68	2.02	2.42	2.70
41	1.30	1.68	2.02	2.42	2.70
42	1.30	1.68	2.02	2.42	2.70
43	1.30	1.68	2.02	2.42	2.70
44	1.30	1.68	2.02	2.41	2.69
45	1.30	1.68	2.01	2.41	2.69
46	1.30	1.68	2.01	2.41	2.69
47	1.30	1.68	2.01	2.41	2.68
48	1.30	1.68	2.01	2.41	2.68
49	1.30	1.68	2.01	2.40	2.68
50	1.30	1.68	2.01	2.40	2.68
60	1.30	1.67	2.00	2.39	2.66

This problem is about a one-sample t-test.

The C.I. is $\bar{x} \pm t^*_{n-1} \cdot SE_{\bar{x}}$.

From the table, $t^*_{32} = 1.69$.

Warm Up

Suppose 33 fish are sampled from a lake. The mean of their lengths is 4.4 inches, and the standard deviation is 1.3 inches. You may assume that the 33 fish represent a random sample of all fish in the lake. Find a 90% confidence interval for the average length of all fish in the lake.

one tail		0.100	0.050	0.025	0.010	0.005
two tails		0.200	0.100	0.050	0.020	0.010
df	31	1.31	1.70	2.04	2.45	2.74
	32	1.31	1.69	2.04	2.45	2.74
	33	1.31	1.69	2.03	2.44	2.73
	34	1.31	1.69	2.03	2.44	2.73
	35	1.31	1.69	2.03	2.44	2.72
	36	1.31	1.69	2.03	2.43	2.72
	37	1.30	1.69	2.03	2.43	2.72
	38	1.30	1.69	2.02	2.43	2.71
	39	1.30	1.68	2.02	2.43	2.71
	40	1.30	1.68	2.02	2.42	2.70
	41	1.30	1.68	2.02	2.42	2.70
	42	1.30	1.68	2.02	2.42	2.70
	43	1.30	1.68	2.02	2.42	2.70
	44	1.30	1.68	2.02	2.41	2.69
	45	1.30	1.68	2.01	2.41	2.69
	46	1.30	1.68	2.01	2.41	2.69
	47	1.30	1.68	2.01	2.41	2.68
	48	1.30	1.68	2.01	2.41	2.68
	49	1.30	1.68	2.01	2.40	2.68
	50	1.30	1.68	2.01	2.40	2.68
	60	1.30	1.67	2.00	2.39	2.66

This problem is about a one-sample t-test.

The C.I. is $\bar{x} \pm t^*_{n-1} \cdot SE_{\bar{x}}$.

From the table, $t^*_{32} = 1.69$.

We know $SE_{\bar{x}} = \dfrac{s_x}{\sqrt{n}} = \dfrac{1.3}{\sqrt{33}} \simeq 0.226$.

Warm Up

Suppose 33 fish are sampled from a lake. The mean of their lengths is 4.4 inches, and the standard deviation is 1.3 inches. You may assume that the 33 fish represent a random sample of all fish in the lake. Find a 90% confidence interval for the average length of all fish in the lake.

one tail		0.100	0.050	0.025	0.010	0.005
two tails		0.200	0.100	0.050	0.020	0.010
df	31	1.31	1.70	2.04	2.45	2.74
	$\boxed{32}$	1.31	1.69	2.04	2.45	2.74
	33	1.31	1.69	2.03	2.44	2.73
	34	1.31	1.69	2.03	2.44	2.73
	35	1.31	1.69	2.03	2.44	2.72
	36	1.31	1.69	2.03	2.43	2.72
	37	1.30	1.69	2.03	2.43	2.72
	38	1.30	1.69	2.02	2.43	2.71
	39	1.30	1.68	2.02	2.43	2.71
	40	1.30	1.68	2.02	2.42	2.70
	41	1.30	1.68	2.02	2.42	2.70
	42	1.30	1.68	2.02	2.42	2.70
	43	1.30	1.68	2.02	2.42	2.70
	44	1.30	1.68	2.02	2.41	2.69
	45	1.30	1.68	2.01	2.41	2.69
	46	1.30	1.68	2.01	2.41	2.69
	47	1.30	1.68	2.01	2.41	2.68
	48	1.30	1.68	2.01	2.41	2.68
	49	1.30	1.68	2.01	2.40	2.68
	50	1.30	1.68	2.01	2.40	2.68
	60	1.30	1.67	2.00	2.39	2.66

This problem is about a one-sample t-test.

The C.I. is $\bar{x} \pm t^*_{n-1} \cdot SE_{\bar{x}}$.

From the table, $t^*_{32} = 1.69$.

We know $SE_{\bar{x}} = \dfrac{s_x}{\sqrt{n}} = \dfrac{1.3}{\sqrt{33}} \simeq 0.226$.

Our interval is

$$4.4 \pm 1.69 \cdot 0.226 = (4.018'', 4.782'').$$

About 90% of the fish in the lake have a length in the confidence interval we just found.

- True
- False

About 90% of the fish in the lake have a length in the confidence interval we just found.

- True
- False

Answer: False
The 90% is about confidence intervals (capturing the true average length), not about fish (having lengths in our one C.I.)

About 90% of the fish in the lake have a length in the confidence interval we just found.

- True
- False

Answer: False
The 90% is about confidence intervals (capturing the true average length), not about fish (having lengths in our one C.I.)

If the level of confidence were increased from 90% to 95%, the width of the interval would increase also.

- True
- False

About 90% of the fish in the lake have a length in the confidence interval we just found.

- True
- False

Answer: False
The 90% is about confidence intervals (capturing the true average length), not about fish (having lengths in our one C.I.)

If the level of confidence were increased from 90% to 95%, the width of the interval would increase also.

- True
- False

Answer: True
The more sure of your guess you want to be, the less precise your guess has to be.
Extreme cases: A 100% C.I. is $(-\infty, +\infty)$. A 0% C.I. is the interval with zero width $(\bar{x}, \bar{x})$.

If we took many samples, about 90% of the sample means would fall in the interval we calculated.

- True
- False

If we took many samples, about 90% of the sample means would fall in the interval we calculated.

- True
- False

Answer: False
The 90% is about confidence intervals capturing the true average length, not about capturing the other possible sample means.

If we took many samples, about 90% of the sample means would fall in the interval we calculated.

- True
- False

Answer: False
The 90% is about confidence intervals capturing the true average length, not about capturing the other possible sample means.

If the sample size is doubled, the width of the 90% C.I. would be cut in half.

- True
- False

If we took many samples, about 90% of the sample means would fall in the interval we calculated.

- True
- False

Answer: False
The 90% is about confidence intervals capturing the true average length, not about capturing the other possible sample means.

If the sample size is doubled, the width of the 90% C.I. would be cut in half.

- True
- False

Answer: False
The width would go from $2 \times t^*_{n-1} \frac{s_x}{\sqrt{n}}$ to $2 \times t^*_{n-1} \frac{s_x}{\sqrt{2n}} = \sqrt{2} \times t^*_{n-1} \frac{s_x}{\sqrt{n}}$.

If we took many samples, about 90% of the confidence intervals we find would contain the average length of all fish in the lake.

- True
- False

If we took many samples, about 90% of the confidence intervals we find would contain the average length of all fish in the lake.

- True
- False

Answer: True
That's the definition of 90% confidence intervals.

As part of a recent College Senior Survey, 685 UCSD seniors (208 men and 477 women) answered about their political views. Of the men, 84 described themselves as liberal, 82 as moderate, and 42 as conservative. Of the women, 250 described themselves as liberal, 163 as moderate, and 64 as conservative.
Do these numbers provide strong evidence of an association between gender and political views among UCSD students? Answer by carrying out an appropriate hypothesis test at significance level 0.05.

As part of a recent College Senior Survey, 685 UCSD seniors (208 men and 477 women) answered about their political views. Of the men, 84 described themselves as liberal, 82 as moderate, and 42 as conservative. Of the women, 250 described themselves as liberal, 163 as moderate, and 64 as conservative.
Do these numbers provide strong evidence of an association between gender and political views among UCSD students? Answer by carrying out an appropriate hypothesis test at significance level 0.05.

First, we get out data organized.

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

As part of a recent College Senior Survey, 685 UCSD seniors (208 men and 477 women) answered about their political views. Of the men, 84 described themselves as liberal, 82 as moderate, and 42 as conservative. Of the women, 250 described themselves as liberal, 163 as moderate, and 64 as conservative.
Do these numbers provide strong evidence of an association between gender and political views among UCSD students? Answer by carrying out an appropriate hypothesis test at significance level 0.05.

First, we get out data organized.

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

This problem asks about an association between two categorical variables (Gender and Political affiliation), so we use a χ^2 test.

As part of a recent College Senior Survey, 685 UCSD seniors (208 men and 477 women) answered about their political views. Of the men, 84 described themselves as liberal, 82 as moderate, and 42 as conservative. Of the women, 250 described themselves as liberal, 163 as moderate, and 64 as conservative.
Do these numbers provide strong evidence of an association between gender and political views among UCSD students? Answer by carrying out an appropriate hypothesis test at significance level 0.05.

First, we get out data organized.

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

This problem asks about an association between two categorical variables (Gender and Political affiliation), so we use a χ^2 test.

We need to calculate the expected cell counts for all six cells. To do so, for each cell we find

$$\frac{\text{Row sum} \cdot \text{Column sum}}{\text{Table sum}}.$$

We get this table of expected counts:

	Liberal	Moderate	Conservative
Men	101.4	74.4	32.2
Women	232.6	170.6	73.8

We get this table of expected counts:

	Liberal	Moderate	Conservative
Men	101.4	74.4	32.2
Women	232.6	170.6	73.8

Recall that the table of observed counts is

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

We get this table of expected counts:

	Liberal	Moderate	Conservative
Men	101.4	74.4	32.2
Women	232.6	170.6	73.8

Recall that the table of observed counts is

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

The conditions for inference are counts, independence of cell counts, and expected cell counts ≥ 5. We might worry about independence because of volunteer bias.

We get this table of expected counts:

	Liberal	Moderate	Conservative
Men	101.4	74.4	32.2
Women	232.6	170.6	73.8

Recall that the table of observed counts is

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

The conditions for inference are counts, independence of cell counts, and expected cell counts ≥ 5. We might worry about independence because of volunteer bias.

$$\begin{aligned}\chi^2 &= \frac{(84-101.4)^2}{101.4} + \frac{(82-74.4)^2}{74.4} + \frac{(42-32.2)^2}{32.2} \\ &\quad + \frac{(250-232.6)^2}{232.6} + \frac{(163-170.6)^2}{170.6} + \frac{(64-73.8)^2}{73.8} \\ &\simeq 9.71.\end{aligned}$$

We get this table of expected counts:

	Liberal	Moderate	Conservative
Men	101.4	74.4	32.2
Women	232.6	170.6	73.8

Recall that the table of observed counts is

	Liberal	Moderate	Conservative	Total
Men	84	82	42	208
Women	250	163	64	477
Total	334	245	106	685

The conditions for inference are counts, independence of cell counts, and expected cell counts ≥ 5. We might worry about independence because of volunteer bias.

$$\begin{aligned}\chi^2 &= \frac{(84-101.4)^2}{101.4} + \frac{(82-74.4)^2}{74.4} + \frac{(42-32.2)^2}{32.2} \\ &\quad + \frac{(250-232.6)^2}{232.6} + \frac{(163-170.6)^2}{170.6} + \frac{(64-73.8)^2}{73.8} \\ &\simeq 9.71.\end{aligned}$$

We have a 3×2 table, so $df = (3-1)\cdot(2-1) = 2$.
We look at the area beyond 9.71 on χ^2_2.

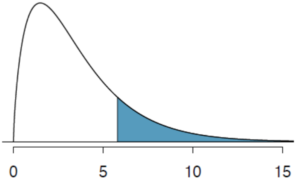


Figure B.2: Areas in the chi-square table always refer to the right tail.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27

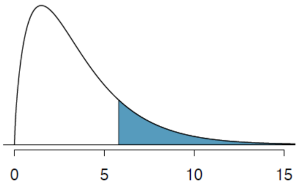


Figure B.2: Areas in the chi-square table always refer to the right tail.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27

9.71 falls between the p-values of 0.01 and 0.005, so $0.005 \leq p \leq 0.01$. Since the range is < 0.05, we reject

H_0: There is no association between gender and political affiliation

in favor of

H_A: There is an association between gender and political affiliation

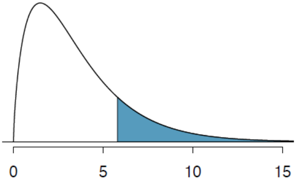


Figure B.2: Areas in the chi-square table always refer to the right tail.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27

9.71 falls between the p-values of 0.01 and 0.005, so $0.005 \leq p \leq 0.01$. Since the range is < 0.05, we reject

H_0: There is no association between gender and political affiliation

in favor of

H_A: There is an association between gender and political affiliation

Notice: You can also say “dependence” instead of “association”.

Which Test Best Fits This Scenario?

You take samples of rain water from 40 different places in a forest, and record the pH level in each sample. You want to determine whether the average pH level of rain in the forest is below 5.6, which is considered to be the threshold for acid rain.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Which Test Best Fits This Scenario?

You take samples of rain water from 40 different places in a forest, and record the pH level in each sample. You want to determine whether the average pH level of rain in the forest is below 5.6, which is considered to be the threshold for acid rain.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Answer: 2. It's a test about a mean in one population

Which Test Best Fits This Scenario?

You want to determine whether, in marriages since the year 2000, the average age of the husband exceeds the average age of the wife by at least 3 years. For 175 randomly chosen couples who were married since the year 2000, you obtain the husband's age and the wife's age.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Which Test Best Fits This Scenario?

You want to determine whether, in marriages since the year 2000, the average age of the husband exceeds the average age of the wife by at least 3 years. For 175 randomly chosen couples who were married since the year 2000, you obtain the husband's age and the wife's age.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Answer: 5. It's a test about a difference of means of two paired populations.

Which Test Best Fits This Scenario?

Out of 200 California voters surveyed, 108 support declaring war to Texas. You want to determine whether at least half of all voters supports declaring war to Texas.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Which Test Best Fits This Scenario?

Out of 200 California voters surveyed, 108 support declaring war to Texas. You want to determine whether at least half of all voters supports declaring war to Texas.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Answer: 1. It's a test about proportions in one population.

Which Test Best Fits This Scenario?

You want to determine whether students at Harvard or Princeton spend more time studying. You survey 50 randomly chosen students at each university and ask them how many hours they spent studying in the last week.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Which Test Best Fits This Scenario?

You want to determine whether students at Harvard or Princeton spend more time studying. You survey 50 randomly chosen students at each university and ask them how many hours they spent studying in the last week.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Answer: 4. It's a test about the difference of means of two independent populations.

Which Test Best Fits This Scenario?

You want to determine whether students at Harvard or Princeton spend more time studying. You survey 50 randomly chosen students at each university and ask them how many hours they spent studying in the last week.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Answer: 4. It's a test about the difference of means of two independent populations.

Notice: To turn this situation into a paired one, you could sample transfert students from one school to the other and survey their personal working time in both schools.

Which Test Best Fits This Scenario?

In a particular town, you know the number of girls in all 236 families that have exactly three children. You want to determine whether the number of girls in such families follows a binomial distribution with $n = 3$ and $p = 0.5$.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Which Test Best Fits This Scenario?

In a particular town, you know the number of girls in all 236 families that have exactly three children. You want to determine whether the number of girls in such families follows a binomial distribution with $n = 3$ and $p = 0.5$.

1. 1-sample z-test
2. 1-sample t-test
3. 2-sample z-test
4. 2-sample t-test
5. Paired t-test
6. χ^2-test in a 1-dimensional table
7. χ^2-test in a 2-dimensional table

Answer: 6. It's a goodness-of-fit test.

Summary of Our Models: Discrete Models

Geometric: $X = Geom(p)$. $X \in \{1, 2, 3, \ldots\}$
X is the number of trials needed to get the first success. Each trial has success probability p.

Binomial: $X = Binom(n, p)$. $X \in \{0, 1, 2, \ldots, n\}$
X is the number of successful trials of out the number of trials. Each trial has success probability p.

Poisson: $X = Poisson(\lambda)$. $X \in \{0, 1, 2, \ldots\}$.
X is the number of times an event occurs in a given times when its average rate of occurrence in that time is λ.

Summary of Our Models: Continuous Models

Uniform: $X = Unif(a, b)$. $$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

X is an output in a finite range for which all outcomes in the range are equally likely.

Exponential: $X = Exp(\lambda)$. $$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

X represents how long we must wait for an event to occur when we know how often it occurs on average. ($\lambda > 0$ is the time-based rate)

Normal: $X = N(\mu, \sigma)$. $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

X represents a real-world phenomenon that is influenced by many independent factors. (μ is the mean and σ the standard deviation)

<u>Recall:</u> A density function satisfies two properties:

$$f(x) \geq 0 \text{ for all values of } x \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

Which Model Best Fits This Situation?

The amount of time before the next volcano eruption on Earth

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The amount of time before the next volcano eruption on Earth

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 5. (Exponential)

Which Model Best Fits This Situation?

The number of emails you will get today.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The number of emails you will get today.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 3. (Poisson)

Which Model Best Fits This Situation?

The number of days we must wait until a politician next tells a lie.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The number of days we must wait until a politician next tells a lie.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 1. (Geometric)

Which Model Best Fits This Situation?

The time of day that a meteor enters Earth's atmosphere.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The time of day that a meteor enters Earth's atmosphere.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 4. (Uniform)

Which Model Best Fits This Situation?

The number of times an archer hits a target when shooting a quiver with ten arrows.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The number of times an archer hits a target when shooting a quiver with ten arrows.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 2. (Binomial)

Which Model Best Fits This Situation?

The length of people's noses

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The length of people's noses

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 6. (Normal)

Which Model Best Fits This Situation?

The number of typos E. Aamari does on one of his lecture slides.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The number of typos E. Aamari does on one of his lecture slides.

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 3. (Poisson)

Which Model Best Fits This Situation?

The number of people the TSA must screen before finding someone who forgot to dispose of his/her water bottle

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The number of people the TSA must screen before finding someone who forgot to dispose of his/her water bottle

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 1. (Geometric)

Which Model Best Fits This Situation?

The average IQ of the three contestants each day on the game jeopardy

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Which Model Best Fits This Situation?

The average IQ of the three contestants each day on the game jeopardy

1. Geometric
2. Binomial
3. Poisson
4. Uniform
5. Exponential
6. Normal
7. t-distribution

Answer: 6. (Normal)

Which Model Best Fits This Situation?

The number of people that get a 5 on AP Stat Exam in a class of 20.

1. Geometric
2. Poisson
3. Binomial
4. Uniform
5. Exponential
6. Normal
7. t-Distribution

Which Model Best Fits This Situation?

The number of people that get a 5 on AP Stat Exam in a class of 20.

1. Geometric
2. Poisson
3. Binomial
4. Uniform
5. Exponential
6. Normal
7. t-Distribution

Answer: 3. (Binomial)
Discrete idea, counting the number of successes, fixed number of "trials" ($n = 20$).

Which Model Best Fits This Situation?

The sampling distribution of the difference in the salaries of UCSD and UCLA graduates in samples of size 200.

1. Geometric
2. Poisson
3. Binomial
4. Uniform
5. Exponential
6. Normal
7. t-Distribution

Which Model Best Fits This Situation?

The sampling distribution of the difference in the salaries of UCSD and UCLA graduates in samples of size 200.

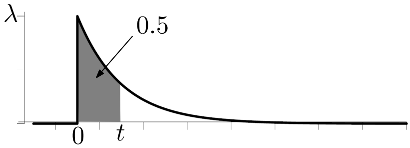
1. Geometric
2. Poisson
3. Binomial
4. Uniform
5. Exponential
6. Normal
7. t-Distribution

Answer: 7. (t-distribution)
It's the sampling distribution of a difference of means $\bar{x}_1 - \bar{x}_2$ for independent samples. Here, $df = \min(200 - 1, 200 - 1) = 199$.

The median of a continuous random variable is the place on the x axis, say t, where 0.5 area is to the right of t under the density function.

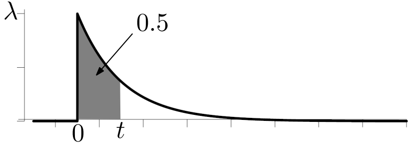
Find the median of the exponential distribution.

The median of a continuous random variable is the place on the x axis, say t, where 0.5 area is to the right of t under the density function.
Find the median of the exponential distribution.



The median of a continuous random variable is the place on the x axis, say t, where 0.5 area is to the right of t under the density function.

Find the median of the exponential distribution.



We want $\int_0^t \lambda e^{-\lambda x} dx = 0.5$.

Thus, $-e^{-\lambda x}\Big|_0^t = 0.5$.

That is, $-e^{-\lambda t} + 1 = 0.5$.

Leading to $0.5 = e^{-\lambda t}$

So, $t = \dfrac{\ln 2}{\lambda}$.

Suppose phone calls arrive at a call center independently of one another at a constant rate of two per minute.
What is the probability that as least two calls arrive in the next minute?

Suppose phone calls arrive at a call center independently of one another at a constant rate of two per minute.
What is the probability that as least two calls arrive in the next minute?

This is asking about a dicrete idea (number of calls) and we are given a rate. We should set $X = Poisson(\lambda = 2)$.

Suppose phone calls arrive at a call center independently of one another at a constant rate of two per minute.
What is the probability that as least two calls arrive in the next minute?

This is asking about a dicrete idea (number of calls) and we are given a rate. We should set $X = Poisson(\lambda = 2)$.

We want

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

Suppose phone calls arrive at a call center independently of one another at a constant rate of two per minute.
What is the probability that as least two calls arrive in the next minute?

This is asking about a dicrete idea (number of calls) and we are given a rate. We should set $X = Poisson(\lambda = 2)$.

We want

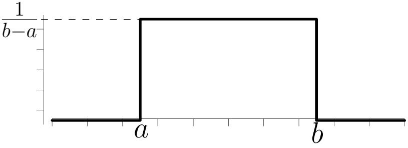
$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

From my sheet of notes, I see $P(X = k) = e^{-\lambda}\lambda^k/k! = e^{-2}2^k/k!$, and hence

$$P(X \geq 2) = 1 - \frac{e^{-2}2^0}{0!} - \frac{e^{-2}2^1}{1!} \simeq 0.594$$

You have seen that 68% of a Normal model is within 1 SD of the mean.
What percent of a uniform distribution is within 1 SD of the mean?

You have seen that 68% of a Normal model is within 1 SD of the mean.
What percent of a uniform distribution is within 1 SD of the mean?

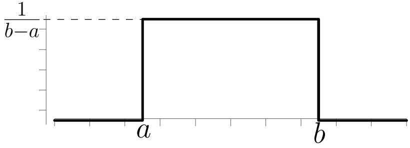


Recall that is $X = Unif(a, b)$,

$$E(X) = \frac{a+b}{2}$$

$$SD(X) = \frac{b-a}{\sqrt{12}}.$$

You have seen that 68% of a Normal model is within 1 SD of the mean.
What percent of a uniform distribution is within 1 SD of the mean?



Recall that is $X = Unif(a, b)$,

$$E(X) = \frac{a+b}{2}$$

$$SD(X) = \frac{b-a}{\sqrt{12}}.$$

If we go 1 SD to each side of the mean, this makes a rectangle with base $\frac{2(b-a)}{\sqrt{12}}$ and height $\frac{1}{b-a}$.

This product is $\frac{2}{\sqrt{12}} \simeq 0.577$.

Suppose that egg weights are normally distributed with a mean of 2 ounces and a standard deviation of 0.2 ounces. A carton of eggs at the store has 12 eggs.
If your goal is to find a carton with a total weight above 25 ounces, how many cartons should you expect to go through before finding one that is heavy enough?
What assumptions did you make in your solution?

Suppose that egg weights are normally distributed with a mean of 2 ounces and a standard deviation of 0.2 ounces. A carton of eggs at the store has 12 eggs.
If your goal is to find a carton with a total weight above 25 ounces, how many cartons should you expect to go through before finding one that is heavy enough?
What assumptions did you make in your solution?

Individual egg weights follow the random variable $W = N(2, 0.2)$. Thus, cartons are modeled by $C = W_1 + \ldots + W_{12}$.

Suppose that egg weights are normally distributed with a mean of 2 ounces and a standard deviation of 0.2 ounces. A carton of eggs at the store has 12 eggs.
If your goal is to find a carton with a total weight above 25 ounces, how many cartons should you expect to go through before finding one that is heavy enough?
What assumptions did you make in your solution?

Individual egg weights follow the random variable $W = N(2, 0.2)$. Thus, cartons are modeled by $C = W_1 + \ldots + W_{12}$.

Assuming that egg weights in a carton are independent, you get

$$E(C) = E(W_1) + \ldots + E(W_{12}) = 12 \times 2,$$

and

$$Var(C) = Var(W_1) + \ldots + Var(W_{12}) = 12 \times 0.2.$$

Therefore,

$$C = N(12 \times 12, \sqrt{12} \times 0.2) = N(24, 0.6928).$$

	Second decimal place					
Z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394

Since $C = N(24, 0.6928)$, writing $Z = \dfrac{C - 24}{0.6928}$, you get

$$
\begin{aligned}
P(C \geq 25) &= P\left(Z \geq \frac{25 - 24}{0.6928}\right) \\
&\simeq P(Z \geq 1.44) \\
&= 1 - P(Z \leq 1.44). \\
&\simeq 1 - 0.9251 \\
&= 0.0759.
\end{aligned}
$$

	Second decimal place					
Z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394

Since $C = N(24, 0.6928)$, writing $Z = \dfrac{C-24}{0.6928}$, you get

$$
\begin{aligned}
P(C \geq 25) &= P\left(Z \geq \frac{25-24}{0.6928}\right) \\
&\simeq P(Z \geq 1.44) \\
&= 1 - P(Z \leq 1.44). \\
&\simeq 1 - 0.9251 \\
&= 0.0759.
\end{aligned}
$$

Let X be the number of carton we examine to find one that is heavy enough.
Then $X = Geom(p = 0.0749)$, and

$$E(X) = \frac{1}{p} \simeq 13.35.$$

You would expect to go through 13.3 cartons before finding one that is heavy enough.

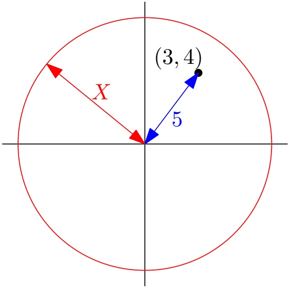
A random value following the exponential distribution $X = Exp(\lambda = 1/2)$ is generated. This value becomes the radius of a circle that is centered at the origin.

1. What is the probability that the point $(3, 4)$ will be inside the circle?

A random value following the exponential distribution $X = Exp(\lambda = 1/2)$ is generated. This value becomes the radius of a circle that is centered at the origin.

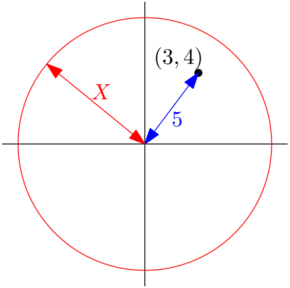
1. What is the probability that the point $(3, 4)$ will be inside the circle?

The circle contains the point $(3, 4)$ if $X > \sqrt{3^2 + 4^2} = 5$.



A random value following the exponential distribution $X = Exp(\lambda = 1/2)$ is generated. This value becomes the radius of a circle that is centered at the origin.

1. What is the probability that the point $(3, 4)$ will be inside the circle?



The circle contains the point $(3, 4)$ if $X > \sqrt{3^2 + 4^2} = 5$.

We want

$$
\begin{aligned}
P(X > 5) &= \int_5^\infty \frac{1}{2} e^{-x/2} dx \\
&= \lim_{n\to\infty} -e^{-x/2}\Big|_5^\infty \\
&= -\lim_{n\to\infty} \left(e^{-n/2} - e^{-5/2}\right) \\
&= e^{-5/2} \\
&=\simeq 0.082.
\end{aligned}
$$

2. Someone states that the expected value for the **area** of the circle should be $\pi \times E(X) \times E(X)$. Explain why this is incorrect and find the correct value.

2. Someone states that the expected value for the **area** of the circle should be $\pi \times E(X) \times E(X)$. Explain why this is incorrect and find the correct value.

The correct answer is $E(\pi X^2) = \pi \times E(X^2) \neq \pi \times E(X) \times E(X)$. Indeed, if $E(X^2)$ and $E(X) \times E(X)$ were equal, then the variance $Var(X) = E(X^2) - E(X)^2$ would be equal to zero.

2. Someone states that the expected value for the **area** of the circle should be $\pi \times E(X) \times E(X)$. Explain why this is incorrect and find the correct value.

The correct answer is $E(\pi X^2) = \pi \times E(X^2) \neq \pi \times E(X) \times E(X)$. Indeed, if $E(X^2)$ and $E(X) \times E(X)$ were equal, then the variance $Var(X) = E(X^2) - E(X)^2$ would be equal to zero.

To find $E(X^2)$, note that $Var(X) = \dfrac{1}{\lambda^2}$, and $E(X) = \dfrac{1}{\lambda}$.
Thus,

$$E(X^2) = Var(X) + E(X)^2 = \frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} = \frac{2}{0.5^2} = 8.$$

Hence,

$$E(\pi X^2) = \pi E(X^2) = 8\pi.$$

The number of books that a store sells per day has a mean of 74 and a standard deviation of 11. The number of magazines that the store sells has a mean of 53 and a standard deviation of 9. Assume the number of books sold is independent of the number of magazines sold. Assume also that the numbers of books and magazines sold on different days are independent of one another.
Calculate the (approximate) probability that the number of books sold in the month of July (31 days) exceeds the number of magazines sold by at least 600.

The number of books that a store sells per day has a mean of 74 and a standard deviation of 11. The number of magazines that the store sells has a mean of 53 and a standard deviation of 9. Assume the number of books sold is independent of the number of magazines sold. Assume also that the numbers of books and magazines sold on different days are independent of one another.
Calculate the (approximate) probability that the number of books sold in the month of July (31 days) exceeds the number of magazines sold by at least 600.

Let X_i be the number of books sold on day i in July.

The number of books that a store sells per day has a mean of 74 and a standard deviation of 11. The number of magazines that the store sells has a mean of 53 and a standard deviation of 9. Assume the number of books sold is independent of the number of magazines sold. Assume also that the numbers of books and magazines sold on different days are independent of one another.
Calculate the (approximate) probability that the number of books sold in the month of July (31 days) exceeds the number of magazines sold by at least 600.

Let X_i be the number of books sold on day i in July.

The number of books sold in July is $X = X_1 + X_2 + \ldots + X_{31}$.

The number of books that a store sells per day has a mean of 74 and a standard deviation of 11. The number of magazines that the store sells has a mean of 53 and a standard deviation of 9. Assume the number of books sold is independent of the number of magazines sold. Assume also that the numbers of books and magazines sold on different days are independent of one another.
Calculate the (approximate) probability that the number of books sold in the month of July (31 days) exceeds the number of magazines sold by at least 600.

Let X_i be the number of books sold on day i in July.

The number of books sold in July is $X = X_1 + X_2 + \ldots + X_{31}$.

We have

$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_{31}) = 31 \cdot 74 = 2294.$$

The number of books that a store sells per day has a mean of 74 and a standard deviation of 11. The number of magazines that the store sells has a mean of 53 and a standard deviation of 9. Assume the number of books sold is independent of the number of magazines sold. Assume also that the numbers of books and magazines sold on different days are independent of one another.
Calculate the (approximate) probability that the number of books sold in the month of July (31 days) exceeds the number of magazines sold by at least 600.

Let X_i be the number of books sold on day i in July.

The number of books sold in July is $X = X_1 + X_2 + \ldots + X_{31}$.

We have

$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_{31}) = 31 \cdot 74 = 2294.$$

By independence, we have

$$Var(X) = Var(X_1) + Var(X_2) + \ldots + Var(X_{31}) = 31 \cdot 11^2 = 3751.$$

Similarly, $Y_1, \ldots, Y_{31}$ is the number of magazines sold on day i.

The number of magazines sold in July is $Y = Y_1 + \ldots + Y_{31}$.

Similarly, $Y_1, \ldots, Y_{31}$ is the number of magazines sold on day i.

The number of magazines sold in July is $Y = Y_1 + \ldots + Y_{31}$.

$$E(Y) = E(Y_1) + \ldots + E(Y_{31}) = 31 \cdot 53 = 1643,$$

and by independence,

$$Var(Y) = Var(Y_1) + \ldots + Var(Y_{31}) = 31 \cdot 9^2 = 2511,$$

Similarly, $Y_1, \ldots, Y_{31}$ is the number of magazines sold on day i.

The number of magazines sold in July is $Y = Y_1 + \ldots + Y_{31}$.

$$E(Y) = E(Y_1) + \ldots + E(Y_{31}) = 31 \cdot 53 = 1643,$$

and by independence,

$$Var(Y) = Var(Y_1) + \ldots + Var(Y_{31}) = 31 \cdot 9^2 = 2511,$$

The problem wants us to compute $P(X - Y \geq 600)$, but

$$E(X - Y) = E(X) - E(Y) = 2294 - 1643 = 651,$$

and by independence,

$$Var(X - Y) = Var(X) + Var(Y) = 3751 + 2511 = 6262.$$

$X - Y$ is a sum of multiple small variations, so it is normal.
Hence, writing $Z = N(0, 1)$, we have

$$
\begin{aligned}
P(X - Y \geq 600) &= P\left(\frac{(X - Y) - 651}{\sqrt{6262}} \geq \frac{600 - 651}{\sqrt{6262}}\right) \\
&\simeq P(Z \geq -0.644) \\
&= P(Z \leq 0.644) \text{ by symmetry.}
\end{aligned}
$$

$X - Y$ is a sum of multiple small variations, so it is normal.
Hence, writing $Z = N(0,1)$, we have

$$P(X - Y \geq 600) = P\left(\frac{(X-Y) - 651}{\sqrt{6262}} \geq \frac{600 - 651}{\sqrt{6262}}\right)$$
$$\simeq P(Z \geq -0.644)$$
$$= P(Z \leq 0.644) \text{ by symmetry.}$$

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

We get:

$$P(X - Y \geq 600) \simeq 73.89\%.$$