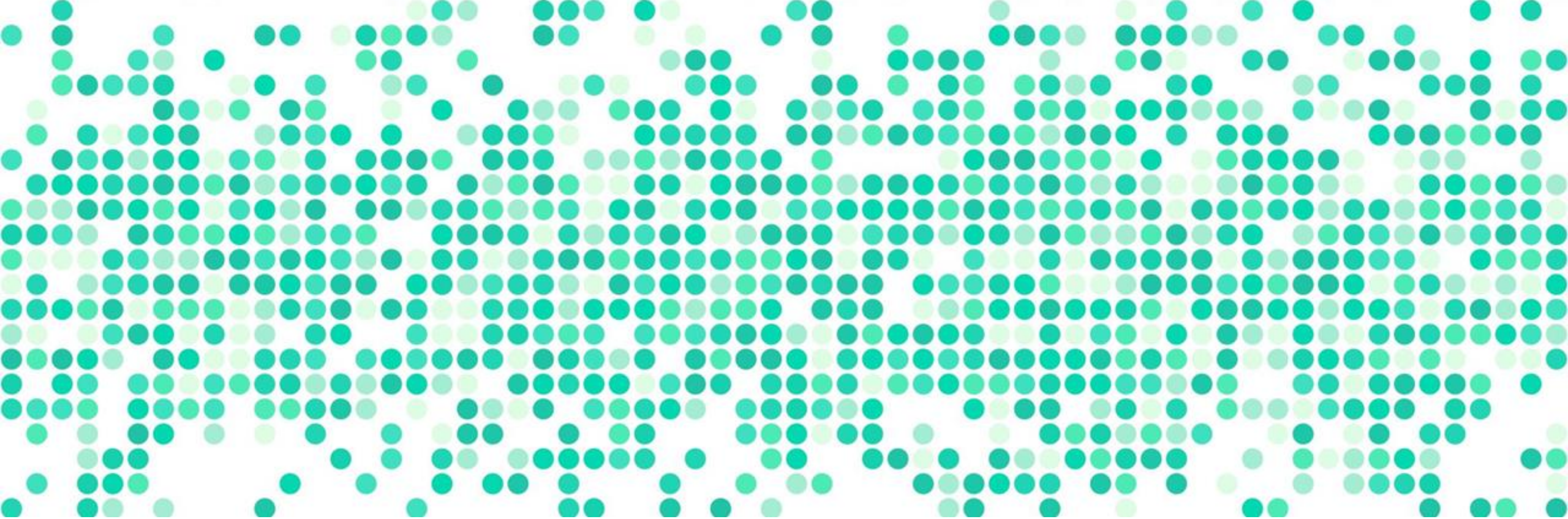
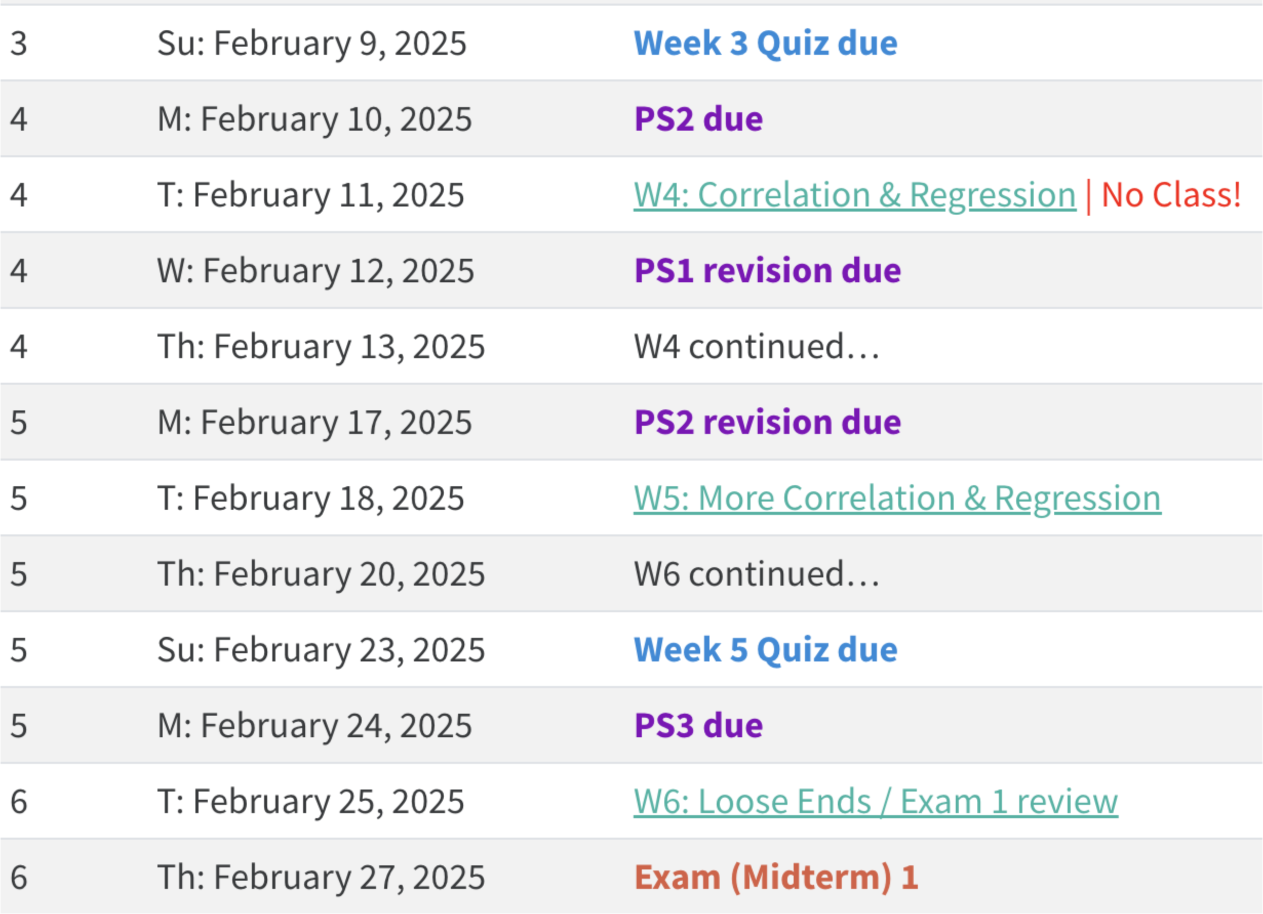


DATA ANALYSIS

Week 3: Normal distribution

logistics: schedule

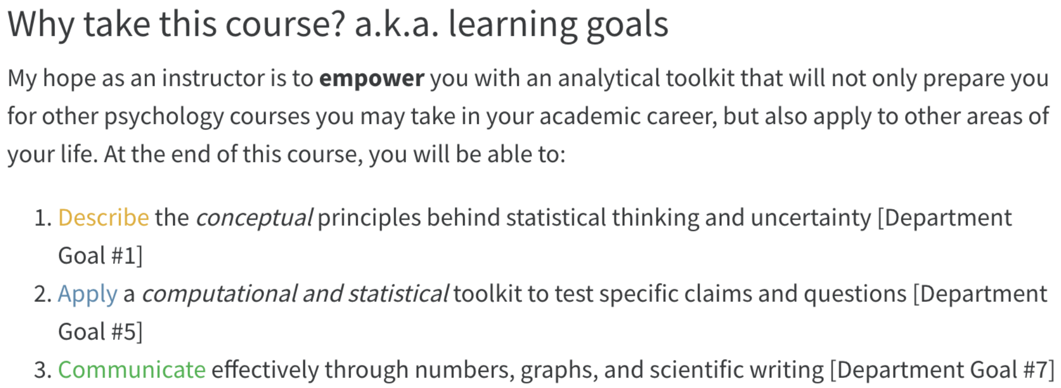
- no office hours today
- no class next Tuesday



3	Su: February 9, 2025	**Week 3 Quiz due**
4	M: February 10, 2025	**PS2 due**
4	T: February 11, 2025	W4: Correlation & Regression | No Class!
4	W: February 12, 2025	**PS1 revision due**
4	Th: February 13, 2025	W4 continued…
5	M: February 17, 2025	**PS2 revision due**
5	T: February 18, 2025	W5: More Correlation & Regression
5	Th: February 20, 2025	W6 continued…
5	Su: February 23, 2025	**Week 5 Quiz due**
5	M: February 24, 2025	**PS3 due**
6	T: February 25, 2025	W6: Loose Ends / Exam 1 review
6	Th: February 27, 2025	**Exam (Midterm) 1**

logistics: problem set 1

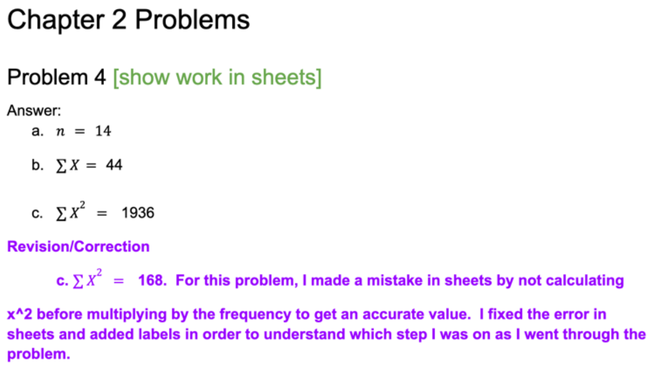
- common mistakes
  - not entering the final answers in the solution sheet for spreadsheet problems
  - not explaining your work
  - incorrect interpretations from frequency tables
  - incorrect cumulative frequencies
  - percentile confusion (also a lingering Q!)
- revisions
  - please explain your work. the goal is not to copy the correct answer
  - I would like to see what you've learned
  - indicate clearly in a different color



Why take this course? a.k.a. learning goals

My hope as an instructor is to **empower** you with an analytical toolkit that will not only prepare you for other psychology courses you may take in your academic career, but also apply to other areas of your life. At the end of this course, you will be able to:

1. Describe the *conceptual* principles behind statistical thinking and uncertainty [Department Goal #1]
2. Apply a *computational and statistical* toolkit to test specific claims and questions [Department Goal #5]
3. Communicate effectively through numbers, graphs, and scientific writing [Department Goal #7]



Chapter 2 Problems

Problem 4 [show work in sheets]

Answer:

a. $n = 14$

b. $\sum X = 44$

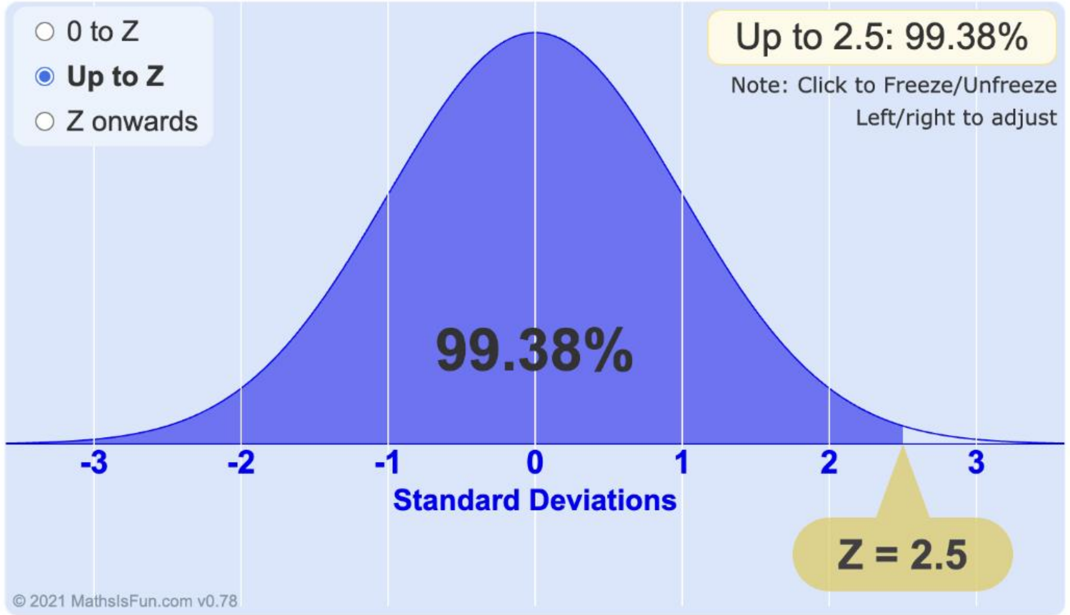
c. $\sum X^2 = 1936$

Revision/Correction

c. $\sum X^2 = 168$. For this problem, I made a mistake in sheets by not calculating x^2 before multiplying by the frequency to get an accurate value. I fixed the error in sheets and added labels in order to understand which step I was on as I went through the problem.

logistics: problem set 2

- you MUST show your work for variance/sd calculations
- you can use VAR.P/STDEV.P and VAR.S/STDEV.S for checking your work, but I need to see all calculations
- for Qs about normal distributions, please provide screenshots from table/website



○ 0 to Z

● Up to Z

○ Z onwards

Up to 2.5: 99.38%

Note: Click to Freeze/Unfreeze
Left/right to adjust

today's agenda

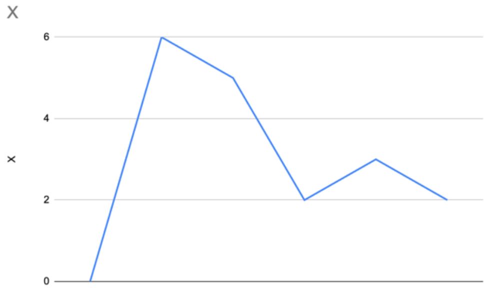
review: z-scores

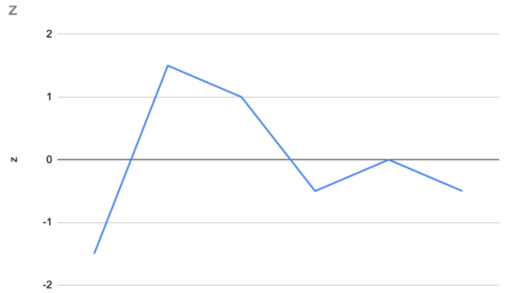
the normal distribution

properties of z-scores

$$z = \frac{X - \mu}{\sigma}$$

- shape of the distribution remains the same before and after z-scoring
- sum of z-scores?
  - always zero! why?
- mean of z-scores?
  - always zero! why?





$$\sum z = \sum \frac{X - \mu}{\sigma} = \frac{1}{\sigma} \sum (X - \mu) = \frac{1}{\sigma}(0) = 0$$

$$M_z = \frac{\sum z}{N} = \frac{0}{N} = 0$$

properties of z-scores

- variance of z-scores?
- always 1! why?

$$z_i = \frac{X_i \; - \; \mu}{\sigma}$$

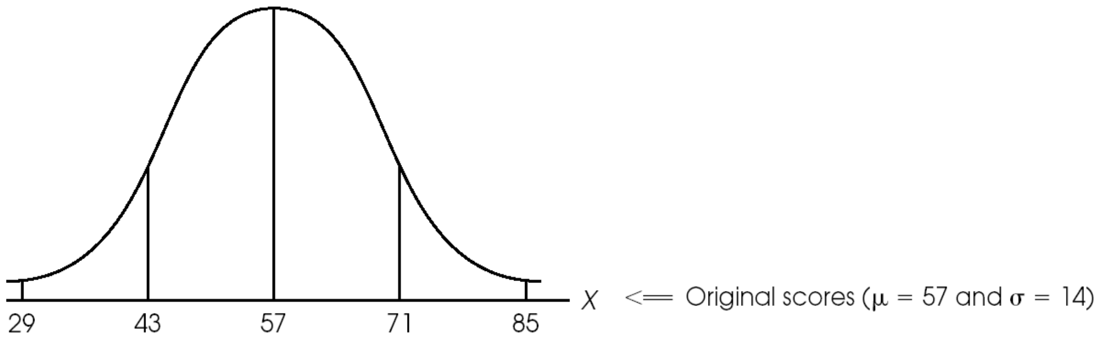
variance of z-scores = $\sigma_z{}^2 = \frac{\sum (z_i - M_z)^2}{N}$

$M_z = 0$, thus $\sigma_z{}^2 = \frac{\sum (z_i)^2}{N} = \frac{\sum (\frac{X \, - \, \mu}{\sigma})^2}{N}$

$= \frac{\frac{1}{\sigma^2} \sum (X - \mu)^2}{N} = \frac{1}{\sigma^2} \; (\sigma^2) = 1$

standardized scores

- z-scoring on original distribution and then obtaining scores on a predetermined μ and σ
- Joe got 43 on original test, where $\mu = 57$ and $\sigma = 14$. What should his score be on a new distribution with $\mu = 50$ and $\sigma = 10$?

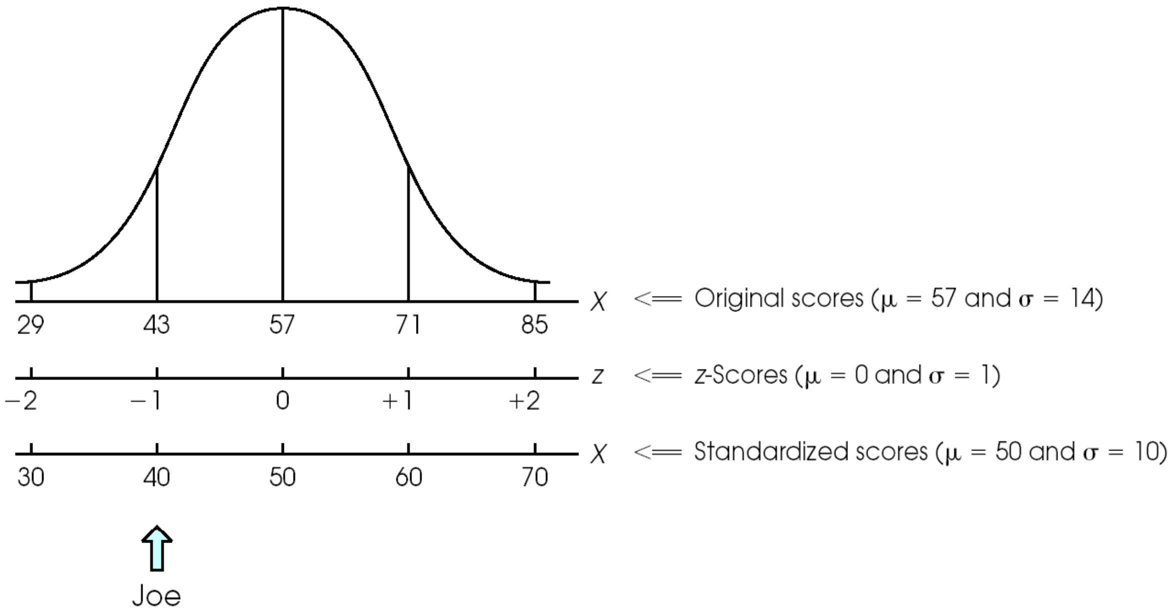


standardized scores

- compute Joe’s z-score on original distribution

$$z = \frac{X\ -\ \mu_1}{\sigma_1} = \frac{43\ \ - 57}{14} = -1$$

- compute Joe’s score on new distribution
  $X =\ \sigma_2 z +\ \mu_2 = 10\ (-1) + 50 = 40$


29
43
57
71
85
X <== Original scores (μ = 57 and σ = 14)
−2
−1
0
+1
+2
z <== z-Scores (μ = 0 and σ = 1)
30
40
50
60
70
X <== Standardized scores (μ = 50 and σ = 10)
Joe

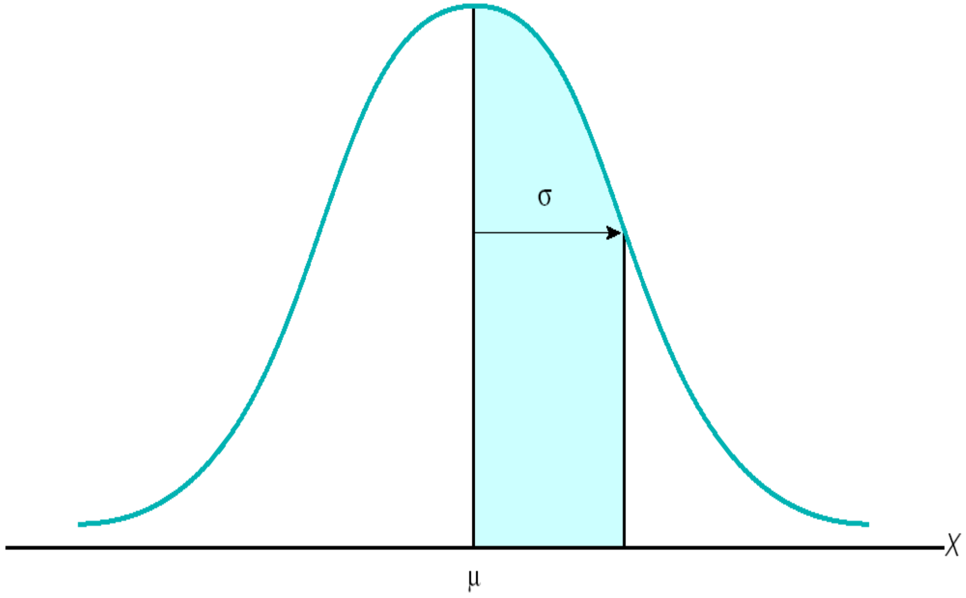
comparing apples and oranges

- Eric competes in two track events: standing long jump and javelin. His long jump is 49 inches, and his javelin throw was 92 ft. He then measures all the other competitors in both events and calculates the mean and standard deviation:
  - Long Jump: M = 44, s = 4
  - Javelin: M = 86ft, s = 10ft
- Which event did Eric do best in?

the normal distribution

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

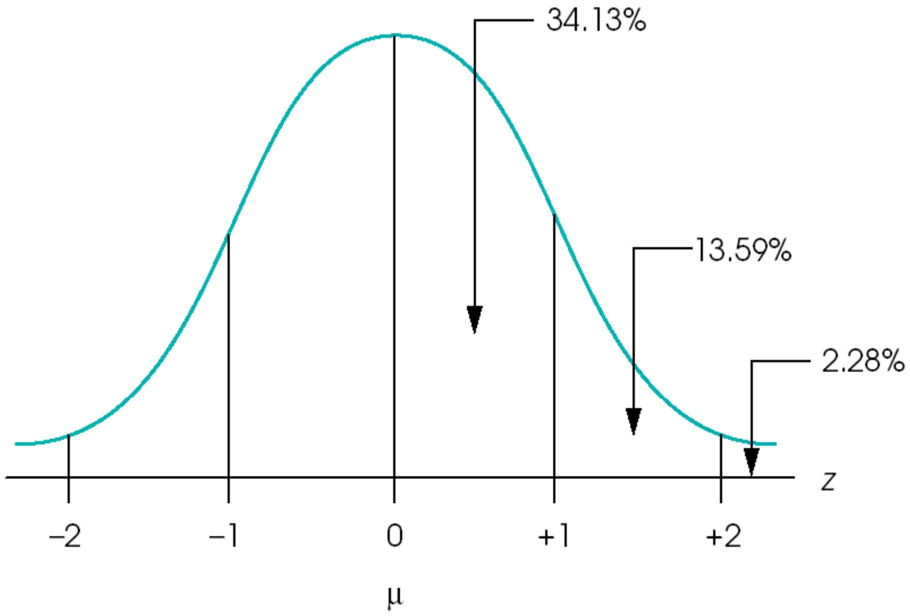
- population distributions typically take the form of a normal distribution
  - symmetric, unimodal, “bell-shaped”
- after converting the normal distribution scores to z-scores, z-scores are often used to identify parts of a normal distribution (μ= 0, σ= 1)
- proportions of areas within the normal distribution can be quantified using z-scores



the normal distribution

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

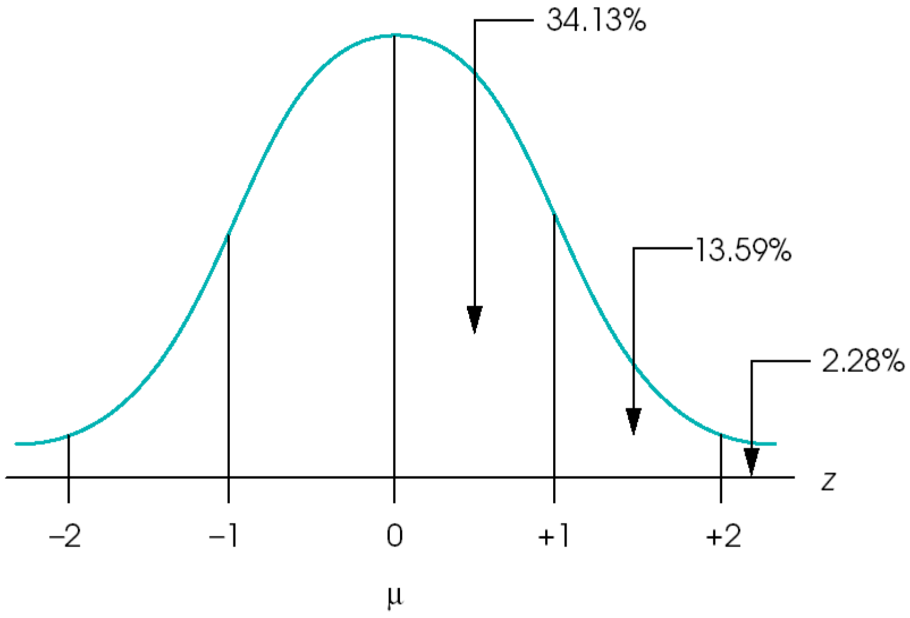
- population distributions typically take the form of a normal distribution
  - symmetric, unimodal, “bell-shaped”
- after converting the normal distribution scores to z-scores, z-scores are often used to identify parts of a normal distribution (μ= 0, σ= 1)
- proportions of areas within the normal distribution can be quantified using z-scores

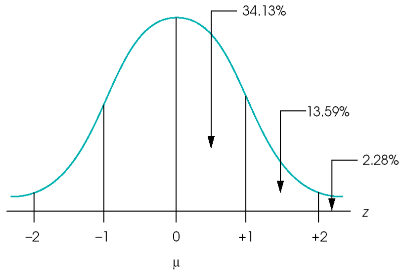


area under the curve

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

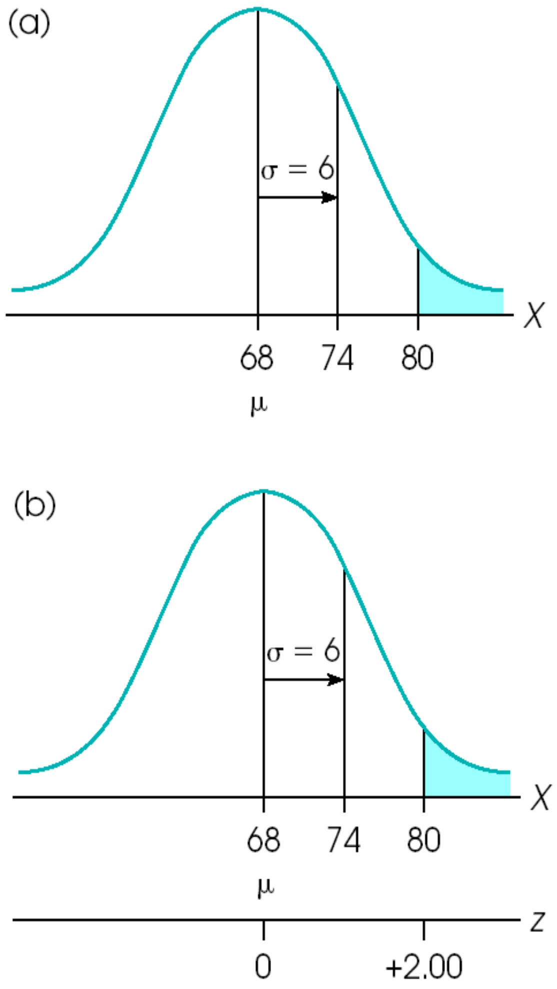
- all normal distributions have the same proportions of data/area in specific locations on the curve
- the normal distribution is symmetrical, i.e., the proportions on both sides of the mean are identical
  - what % of the scores are above the mean?
  - what % of the scores are above 2 standard deviations?
- ~ 68% of scores fall between z-scores of -1 and +1
- ~ 95% fall between z-scores of -2 and +2
- ~ 99% fall between z-scores of -3 and +3



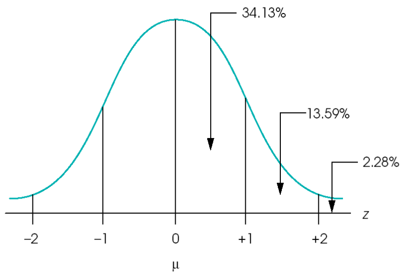


example

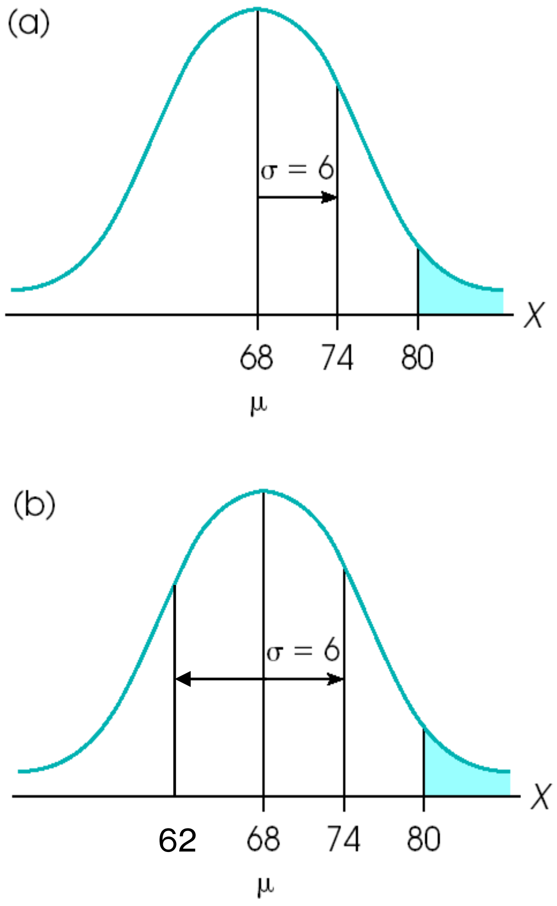
- body height has a normal distribution, with μ= 68, and σ= 6.
- if we select one person at random, what is the probability for selecting a person taller than 80?
  - represent the problem graphically
  - convert to z-scores, 80 is 2 σ
  - all scores **above** 2 $\sigma = 2.28\%$



example

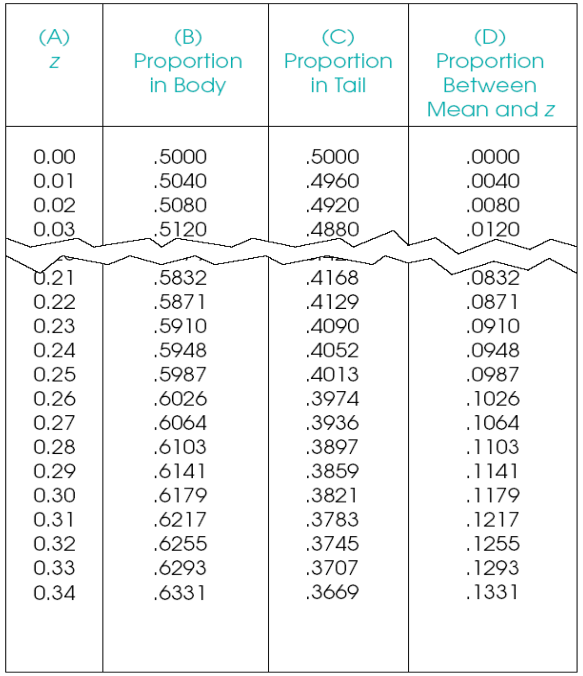


- body height has a normal distribution, with μ= 68, and σ= 6.
- what % of people have heights between 62 and 74?
  - represent the problem graphically
  - convert to z-scores, 62 is -1σ and 74 is 62 is +1σ
    - all scores **<u>between</u>** -σ to +σ
    - 34.13 + 34.13 = 68.26

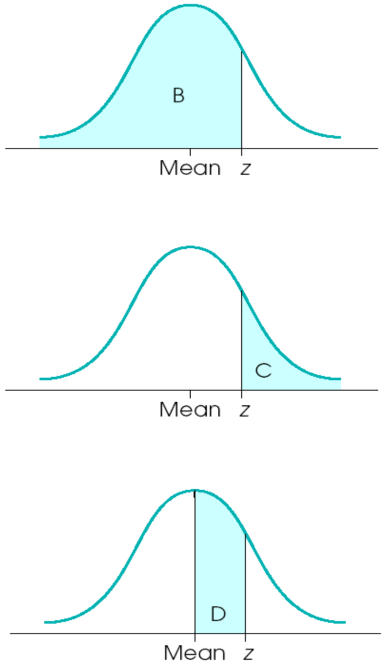


unit normal table / calculator

- questions about whole z-scores (±1, ±2, etc.) are easily gleaned from the distribution, but estimates for fractional z-scores are trickier to obtain via eyeballing
- unit normal tables contain proportion estimates for the full scale of possible z-scores
  - column A: the z-score (vertical line)
  - column B (body): the larger section created by the z-score
  - column C (tail): smaller section created by z-score
  - column D: section between mean and z-score
- available in several places online!
  - full table
  - visual calculator



(A) *z*	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and *z*
0.00	.5000	.5000	.0000
0.01	.5040	.4960	.0040
0.02	.5080	.4920	.0080
0.03	.5120	.4880	.0120
0.21	.5832	.4168	.0832
0.22	.5871	.4129	.0871
0.23	.5910	.4090	.0910
0.24	.5948	.4052	.0948
0.25	.5987	.4013	.0987
0.26	.6026	.3974	.1026
0.27	.6064	.3936	.1064
0.28	.6103	.3897	.1103
0.29	.6141	.3859	.1141
0.30	.6179	.3821	.1179
0.31	.6217	.3783	.1217
0.32	.6255	.3745	.1255
0.33	.6293	.3707	.1293
0.34	.6331	.3669	.1331

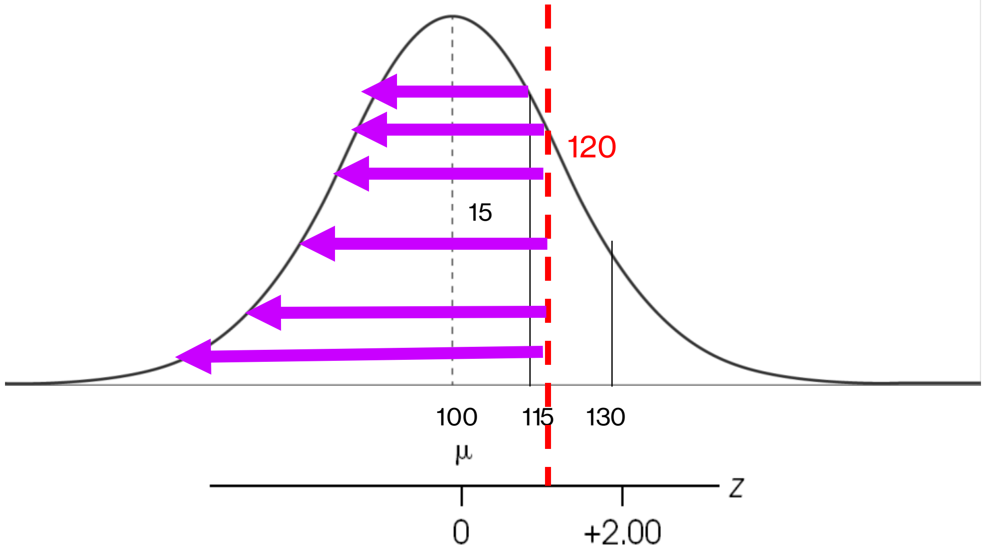


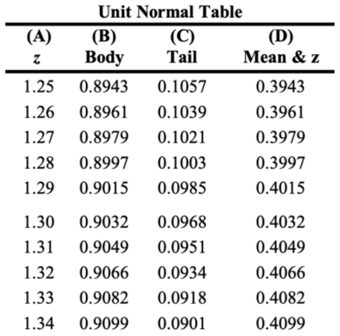
probabilities from scores: example 1

- for an IQ test, the known population parameters are μ= 100, and σ= 15. What is the probability of randomly selecting an individual with an IQ score of **less than 120**?
- represent the problem graphically
- transform X into z
- look up full table (column B) or visual calculator

probabilities from scores: example 1

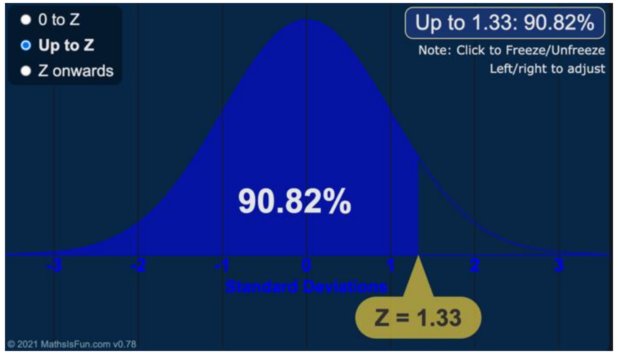
- for an IQ test, the known population parameters are μ= 100, and σ= 15. What is the probability of randomly selecting an individual with an IQ score of **<u>less than 120</u>**?
- represent the problem graphically
- transform X into z = $\frac{X-\mu}{\sigma} = \frac{120-100}{15} = 1.33$
- full table (column B) or visual calculator
- proportion or probability = .9082
- percentage = 90.82%





Unit Normal Table

(A) z	(B) Body	(C) Tail	(D) Mean & z
1.25	0.8943	0.1057	0.3943
1.26	0.8961	0.1039	0.3961
1.27	0.8979	0.1021	0.3979
1.28	0.8997	0.1003	0.3997
1.29	0.9015	0.0985	0.4015
1.30	0.9032	0.0968	0.4032
1.31	0.9049	0.0951	0.4049
1.32	0.9066	0.0934	0.4066
1.33	0.9082	0.0918	0.4082
1.34	0.9099	0.0901	0.4099



○ 0 to Z
● Up to Z
○ Z onwards

Up to 1.33: 90.82%

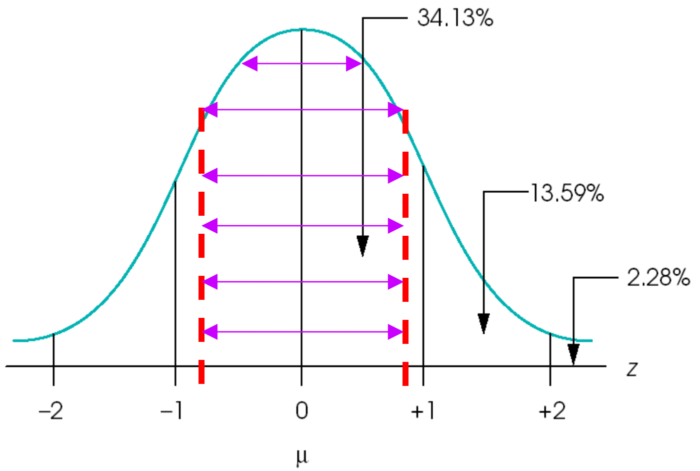
Note: Click to Freeze/Unfreeze
Left/right to adjust

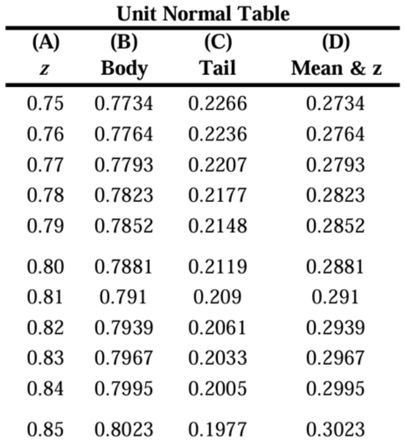
z-scores from proportions: example 2

- what z-score values form the boundaries that separate the middle 60% of the distribution?
- represent the problem graphically
- transform X into z
- full table (column D) or visual calculator

z-scores from proportions: example 2

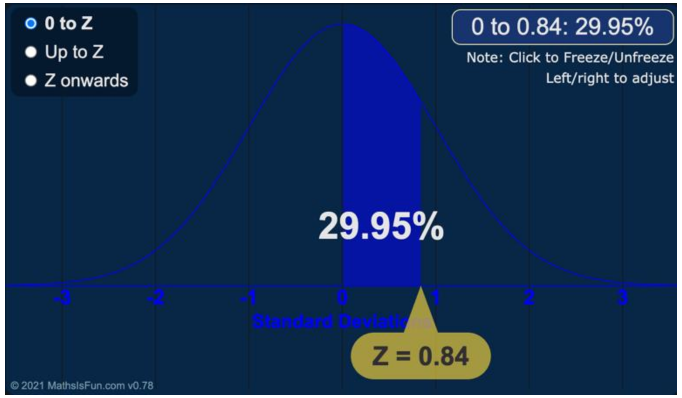
- what z-score values form the boundaries that separate the middle 60% of the distribution?
- represent the problem graphically
  - somewhere less than 34% on both sides
- transform X into z
  - find z-value for p close to ±0.30
- full table (column D) or visual calculator
- z = -0.84 to +0.84





Unit Normal Table

(A) *z*	(B) Body	(C) Tail	(D) Mean & *z*
0.75	0.7734	0.2266	0.2734
0.76	0.7764	0.2236	0.2764
0.77	0.7793	0.2207	0.2793
0.78	0.7823	0.2177	0.2823
0.79	0.7852	0.2148	0.2852
0.80	0.7881	0.2119	0.2881
0.81	0.791	0.209	0.291
0.82	0.7939	0.2061	0.2939
0.83	0.7967	0.2033	0.2967
0.84	0.7995	0.2005	0.2995
0.85	0.8023	0.1977	0.3023



○ 0 to Z
● Up to Z
● Z onwards

0 to 0.84: 29.95%

Note: Click to Freeze/Unfreeze
Left/right to adjust

© 2021 MathsIsFun.com v0.78

W3 Activity 3

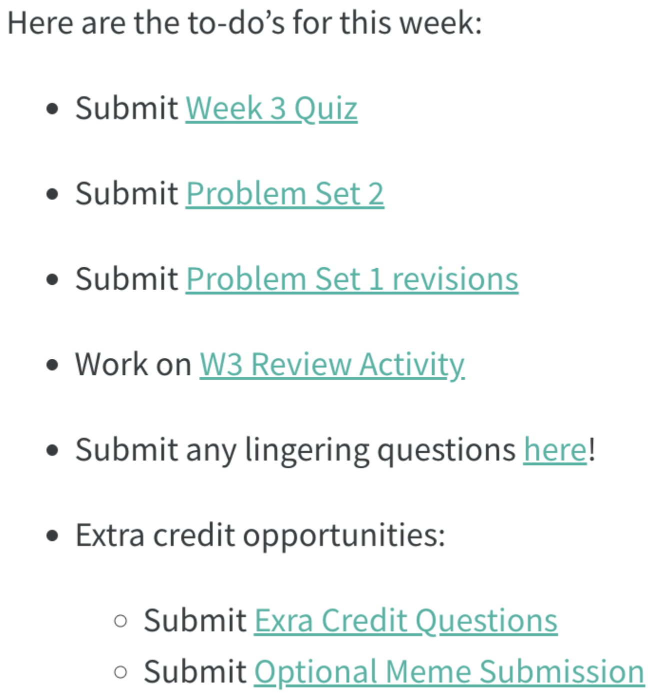
- complete the activity on your own
- discuss the logic behind your answers with a peer
- re-attempt the activity

W3-W4 review activity

- start reviewing concepts and preparing for midterm
- the format of these questions mirror the format of the computational exam
- answer the questions
- answer sheet will be released in Week 6 (Review before Exam 1)

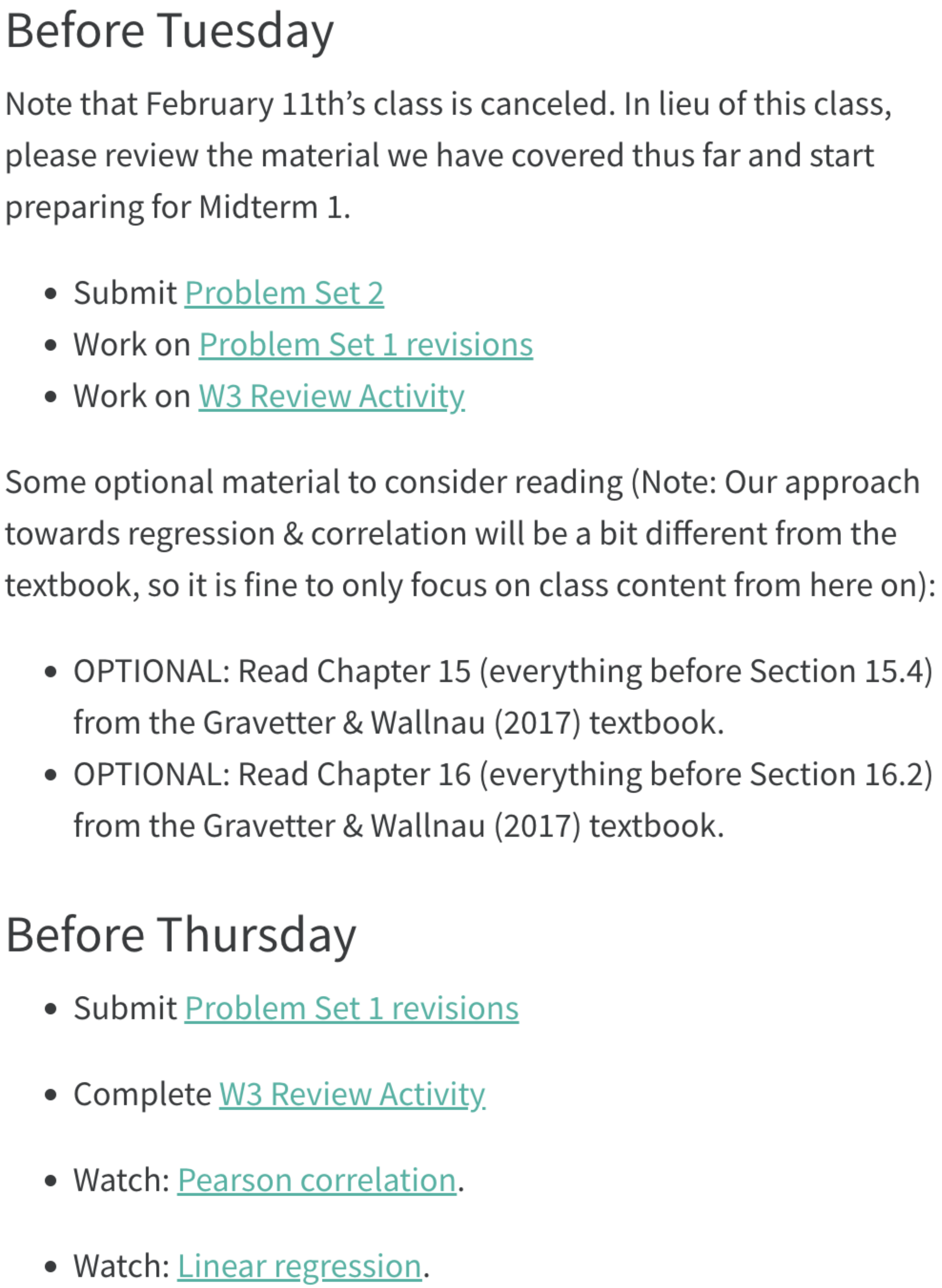
next time

- **no class on Feb 11 (Tuesday)**



Here are the to-do's for this week:

- Submit Week 3 Quiz
- Submit Problem Set 2
- Submit Problem Set 1 revisions
- Work on W3 Review Activity
- Submit any lingering questions here!
- Extra credit opportunities:
  - Submit Exra Credit Questions
  - Submit Optional Meme Submission



Before Tuesday

Note that February 11th's class is canceled. In lieu of this class, please review the material we have covered thus far and start preparing for Midterm 1.

- Submit Problem Set 2
- Work on Problem Set 1 revisions
- Work on W3 Review Activity

Some optional material to consider reading (Note: Our approach towards regression & correlation will be a bit different from the textbook, so it is fine to only focus on class content from here on):

- OPTIONAL: Read Chapter 15 (everything before Section 15.4) from the Gravetter & Wallnau (2017) textbook.
- OPTIONAL: Read Chapter 16 (everything before Section 16.2) from the Gravetter & Wallnau (2017) textbook.

Before Thursday

- Submit Problem Set 1 revisions
- Complete W3 Review Activity
- Watch: Pearson correlation.
- Watch: Linear regression.