STATS BOOTCAMP PRACTICE PROBLEMS

1. Variable types, central tendency, frequency distributions, and st. dev.

1. Education is what type of variable if measured as number of years of schooling?

ratio

2. Income is what type of variable if people are classified as "low income," "middle income," or "high income"?

ordinal

3. We surveyed 30 UCI students to see how many movies they went to see in the past week (X). Results are below. What are the relative frequencies of movies seen? What are the mean/median/mode?

X	Frequency	Rel. Freq.	Freq(x)	Cum. Freq.
0	14	.467	0	14
1	6	.2	6	20
2	8	.267	16	28
3	1	.033	3	29
4	1	.033	4	30
	30		29	

mean $= \frac{29}{30} = 0.967$

median = 1

mode = 0

4. Each person in a sample of 10 residing in Southern California are asked their area code. The responses are:

714, 562, 562, 858, 858, 858, 714, 310, 323, 562

(a) For this data set, construct a frequency and relative frequency distribution.

X	Freq.	Rel. Freq
714	2	.2
562	3	.3
858	3	.3
310	1	.1
323	1	.1
	10	

(b) Is this variable nominal, ordinal or interval?

nominal

5. UCI students recorded the number of magazines they read last month. The results are shown below:

Student	# Magazines (x)	$x-\bar{x}$	$(x-\bar{x})^2$
1	2	-1	1
2	1	0	0
3	3	1	1
4	2	2	4
5	4	3	9

What are the mean and standard deviation the number magazines read?

$n = 5$

$\Sigma x = 10$

$\bar{x} = \frac{10}{5} = 2$

$\Sigma (x-\bar{x})^2 = 15$

$\sigma = \sqrt{\frac{15}{5}} = 1.73$

2. Probability, Z scores, and Sampling Distributions

1. I have a standard deck of 52 playing cards. What is the probability that I randomly pull a red card? A jack? The jack of diamonds?

red $= \frac{13}{52} = 0.25$

jack $= \frac{4}{52} = 0.0769$

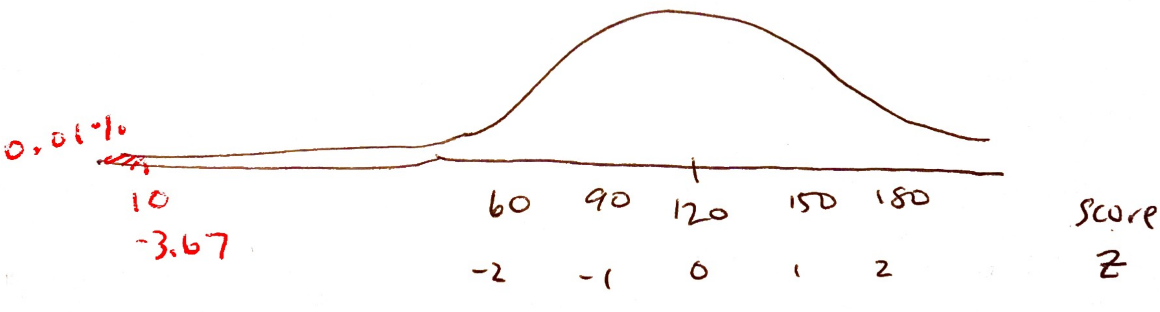
jack of diamonds $: \frac{1}{52} = 0.0192$

2. Fill out the rest of the table with the missing probabilities (expressed as percentages) and Z scores.

* these are left tail probabilities

Z score	p-value
-2.91	0.18%
-0.06	47.61%
1.57	94.18%
3.13	99.91%
-2.45	0.71%
-1.18	11.90%
2.00	97.72%
2.88	99.80%

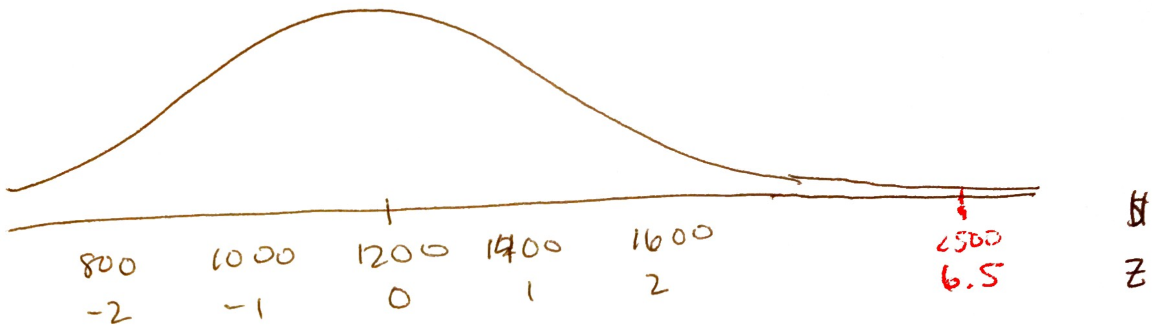
3. Given a set of normally distributed observations with mean of 120 and standard deviation of 30, what is the probability of randomly choosing a number less than 10?



$$z = \frac{10 - 120}{30} = -3.67$$

$p = 0.0001 \approx 0$

4. Data from a large survey shows that on average, people spend $1,200 on travel throughout the year. Respondent A, however, spends $2,500 on travel. Given a standard deviation of $200, and assuming the data is normally distributed, how many standard deviations from the mean is Respondent A? (AKA what is the z-score for this respondent?)



$$z = \frac{2500 - 1200}{200} = 6.5$$

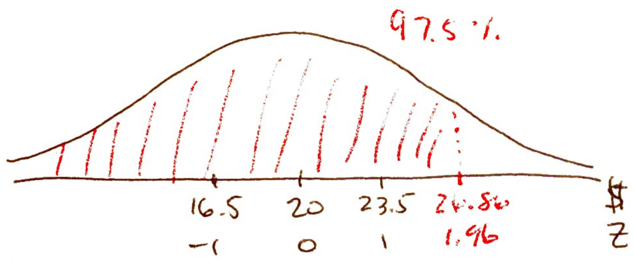
5. Assume hourly wages in a company are normally distributed with a mean of $20 and a standard deviation of $3.50. Compute the following:

(a) The 97.5th percentile of hourly wages.

p = .9750

z = 1.96

x = 20 + 1.96 (3.5) = $26.86

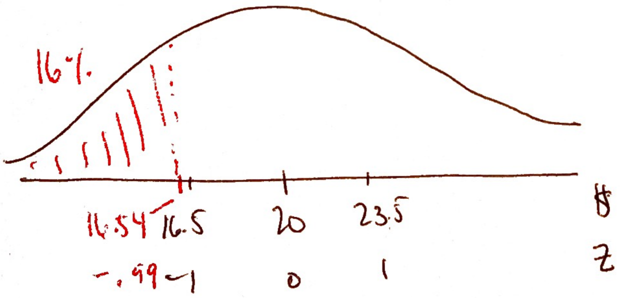


(b) The 16th percentile of hourly wages.

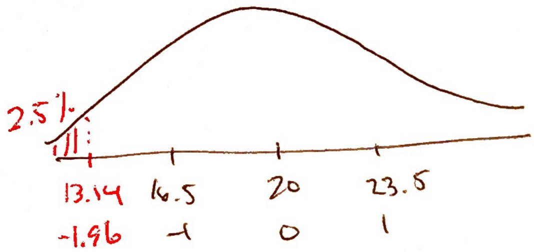
p = .1600

z = -0.99

x = 20 − .99 (3.5) = $16.54



(c) What hourly wage could be used to distinguish between the lowest 2.5% of wage earners and those with higher wages?

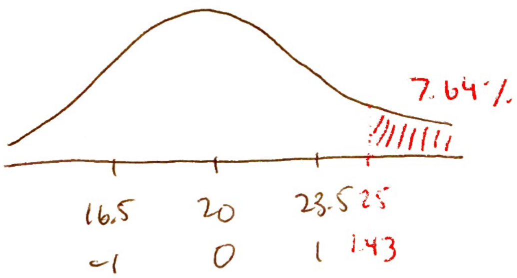


$p = .0250$

$z = -1.96$

$x = 20 - 1.96(3.5) = \$13.14$

(d) What is the probability that, selecting an employee at random, you would pick someone who made more than $25?



$x = 25$

$z = \frac{25-20}{3.5} = 1.43$

$p = 1 - .9236 = .0764$

3. Confidence Intervals

1. You interview 200 people about their income, and calculate a mean income of \$35,000, with a standard deviation of \$1,000. Construct a 95% confidence interval for the population mean income.

$\bar{X} = 35000 \quad s = 1000 \quad n = 200$

$95\% \longrightarrow z = 1.96$

$35000 \pm 1.96\left(\frac{1000}{\sqrt{200}}\right)$

35000 ± 138.59

$(34861.41,\ 35138.59)$

2. After asking 10 students how much tv they watched during the Olympics, you get a mean of 12 hours per week with a standard deviation of 3. With 90% confidence, what is the average amount of tv watched by all UCI students?

$\bar{x} = 12 \quad s = 3 \quad n = 10$

$df = 9,\ 90\% \longrightarrow t = 1.833$

$12 \pm 1.833\left(\frac{3}{\sqrt{10}}\right)$

12 ± 1.74

$(10.26,\ 13.74)$

3. A sample of 84 UCI students were asked how many units they?re currently taking. The sample data are as follows:

Units	Frequency
8	15
12	36
14	21
16	12

What proportion of students in the sample are taking 14 units? Construct a 95% confidence interval for the proportion of students out of the whole sample taking 14 units.

$$\hat{\pi} = \frac{21}{84} = .25$$

$$\sigma = \sqrt{.25(1-.25)} = .43$$

$$\sigma_{\hat{\pi}} = \frac{.43}{\sqrt{84}} = .047$$

$$.25 \pm 1.96(.047)$$

$$.25 \pm .093$$

$$(.157, .343)$$

4. What proportion of students use the library on campus? After asking 150 people, you find out that 100 of them regularly use the library. Based on a 95% confidence interval, use these figures to estimate the proportion of students that regularly go to the library.

$$\hat{\pi} = \frac{100}{150} = .67$$

$$\sigma = \sqrt{.67(1-.67)} = .47$$

$$\sigma_{\hat{\pi}} = \frac{.47}{\sqrt{150}} = .038$$

$$.67 \pm 1.96(.038)$$

$$.67 \pm .075$$

$$(.591, .742)$$

4. Hypothesis Testing

1. The difference between milk chocolate and dark chocolate is the percentage of cacao in the chocolate (dark chocolate has a higher cacao content, while milk chocolate has less cacao). Do chocolate-eaters prefer prefer cacao concentrations of more than 60%?

$H_0: \mu = 60\%$ $\quad H_A: \mu > 60\%$

(a) We survey 130 people and found that the average desired percentage of cacao is 63.4%, with a standard deviation of 17%. Is this significantly higher than 60%, at $\alpha = .05$?

$\bar{x} = 63.4 \quad s = 17 \quad \mu_0 = 60 \quad n = 130$

$$t = \frac{63.4 - 60}{17/\sqrt{130}} = 2.28$$

$p = .0113 < \alpha = .05$

reject H_0

63.4% is significantly higher than 60% at $\alpha = .05$.

(b) What if our sample size had been 25?

$\bar{x} = 63.4 \quad s = 17 \quad \mu_0 = 60 \quad n = 25$

$$t = \frac{63.4 - 60}{17/\sqrt{25}} = 1$$

$p = .163 > \alpha = .05$

fail to reject H_0

63.4% is not significantly higher than 60% at $\alpha = .05$.

2. Research suggests that sociology grad students like free food. In fact, it?s reported that 80% of them will attend a one-hour lecture on proper ASA citation format if there will be food served. We surveyed 50 students and found that only 60% would attend this lecture. Perform a hypothesis test with $\alpha = 0.05$ to test whether the proportion of students who want the free food is less than 80%.

$H_0 : \pi = .8 \qquad H_A : \pi < .8$

Assumptions

$50(.6) = 30 > 5$ ✓

$50(.4) = 20 > 5$ ✓

$\hat{\pi} = .6 \qquad \sigma = \sqrt{.6(1-.6)} = .49 \qquad \sigma_{\hat{\pi}} = \frac{.49}{\sqrt{50}} = .07$

$$z = \frac{.6 - .8}{.07} = -10.8$$

$p \approx 0 < \alpha = .05$

reject H_0

The proportion of students who want free food is significantly less than 80% at $\alpha = .05$.

3. Your roommate claims that a whole bunch of coyotes lives nearby. So many, that in any given week you will probably have 8 coyote sightings. You decide to test this by looking for coyotes around campus. After 25 trials, you find, on average, only 6 sightings in a given week (standard deviation = 5).

(a) Is your roommate right, or are there fewer coyotes than they claim? Test their hypothesis with $\alpha = .05$.

$H_0 : \mu = 8 \qquad H_A : \mu < 8$

$\bar{x} = 6 \qquad s = 5 \qquad \mu_0 = 8 \qquad n = 25$

$$t = \frac{6-8}{\sqrt{25}} = -0.4$$

$p = 0.346 > \alpha = .05$

fail to reject H_0

there is not enough evidence to suggest there will be fewer than 8 coyote sightings in a given week, at $\alpha = .05$.

(b) What is a probable answer for the average number of coyote sightings?

8

4. Research suggests that people who like Batman are more likely to read a lot of comics than people who like Superman. After you collect some data, you find the following information:

	(1) Batman Fans	(2) Superman Fans
Mean # comics read last year	60	45
Standard deviation	10	12
n	100	100

Test whether this hypothesis is true at $\alpha = .05$.

$H_0: \mu_1 = \mu_2 \qquad \mu_1 - \mu_2 = 0$

$H_A: \mu_1 > \mu_2 \qquad \mu_1 - \mu_2 > 0$

$\bar{x}_1 = 60 \qquad s_1 = 10 \qquad n_1 = 100$

$\bar{x}_2 = 45 \qquad s_2 = 12 \qquad n_2 = 100$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{10^2}{99} + \frac{12^2}{99}} = 1.57$$

$$t = \frac{60-45}{1.57} = 9.55$$

$p \approx 0 < \alpha = .05$

reject H_0

Batman fans read significantly more comics than Superman fans at $\alpha = .05$.

5. It's a fact: graduate students live longer than non-graduate students. You want to test this, and so collect some data on education and living past 90:

		Lived past 90	Did not live past 90	Row total
(1)	Went to grad school	10	28	38
(2)	Did not go to grad school	12	50	62
	Column total	22	78	100

Do those who attended graduate school live longer (i.e. are the proportion of graduates who live past 90 larger than the proportion of non-graduates who live past 90)? Test this hypothesis at the $\alpha = .05$ level.

$$H_0: \pi_1 = \pi_2 \qquad \pi_1 - \pi_2 = 0$$

$$H_A: \pi_1 > \pi_2 \qquad \pi_1 - \pi_2 > 0$$

$$\hat{\pi}_1 = \frac{10}{38} = .263$$

$$\hat{\pi}_2 = \frac{12}{62} = .194$$

$$\hat{\pi} = \frac{38(.263) + 62(.194)}{38+62} = \frac{10+12}{100} = .22$$

$$\sigma_{\hat{\pi}_1 - \hat{\pi}_2} = \sqrt{\frac{.22(1-.22)}{38} + \frac{.22(1-.22)}{62}} = .0853$$

$$z = \frac{.263 - .194}{.0853} = 5.35$$

$$p \approx 0 < \alpha = .05$$

reject H_0

Evidence suggests that graduate students live significantly longer than non-graduate students at $\alpha = .05$.