Math 214 Statistics Spring 19 ~~Sample~~ Midterm 2

Name: Solutions

- I will count your best 8 of the following 10 questions.

1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
9	10	
10	10	
	80	

Midterm 2	
Overall	

1. You take a sample of size $n = 10,000$ from a population with mean $\mu = 12$ and standard deviation $\sigma = 20$. Indicate whether or not the following statements are true or false.

(a) The mean of the sample $\overline{x}$ should be close to 12.

T

⊗ True
○ False

(b) The standard deviation s of the sample should be close to 20.

T

⊗ True
○ False

(c) The standard deviation s of the sample should be close to $20/\sqrt{10,000}$.

F

○ True
⊗ False

(d) The sample should be approximately normally distributed.

F

○ True
⊗ False

(e) The distribution of the sample should be close to the distribution of the original population.

T

⊗ True
○ False

2. A researcher chooses 30 New Yorkers from the ages of 20–29 and asks them how many hot dogs they have eaten in the past month. The average number of hot dogs eaten is 8.2. Assume that you know that the overall population standard deviation is $\sigma = 6.3$.

   (a) What is the distribution of the sample mean $\overline{x}$?

   (b) Find a 98% confidence interval for the population mean.

4 a) $\bar{x} \underset{\text{approx}}{\sim} N\left(\mu, \frac{6.3}{\sqrt{30}}\right)$

6 b) $\bar{x} \pm z^* \frac{6.3}{\sqrt{30}}$, where $z^* = 2.33$

$8.2 \pm 2.33 \cdot \frac{6.3}{\sqrt{30}} = (5.52, 10.88)$

3. Consider the previous example, but more realistically, assume you do not know the population standard deviation, but you do know that the sample standard deviation is $s = 6.3$.

   (a) What is the distribution of the sample mean $\overline{x}$?

   (b) Find a 90% confidence interval for the population mean.

4 a) $\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim$ t-dist w/ 29 degrees of freedom t_{29}, so $\bar{x} \sim t_{29} \cdot \frac{s}{\sqrt{n}} + \mu$.

6 b) $\bar{x} \pm t^*_{29} \frac{6.3}{\sqrt{30}}$ $\qquad t^*_{29} = 1.699$

$8.2 \pm 1.699 \times \frac{6.3}{\sqrt{30}} = (6.25, 10.15)$

4. It is known that the average number of hot dogs consumed each month by people in the US is 5.7. Consider the sample from the previous question, and use it to investigate whether New Yorkers eat significantly more or less hot dogs than the national average. Choose the confidence level to be $C = 95\%$.

(a) Set up a null hypothesis, and an alternative hypothesis.

(b) Describe the test statistic, and compute it.

(c) What is the p-value for the test statistic?

(d) Draw a conclusion.

2 a) $H_0: \mu = 5.7$

$H_a: \mu \neq 5.7$

3 b) $\frac{\bar{x} - 5.7}{6.3/\sqrt{30}} \sim t_{29}$ $\qquad \frac{8.2 - 5.7}{6.3/\sqrt{30}} = 2.174$

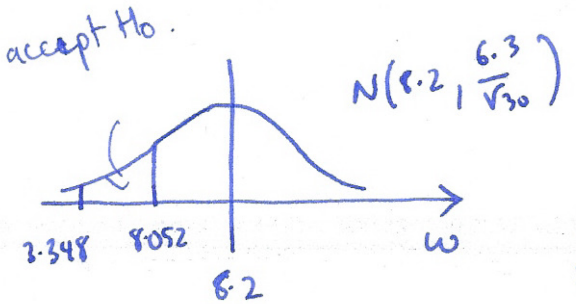
3 c) $\cancel{0.96}\ 0.02 \leq \text{p-value} \leq \cancel{0.98}\ 0.04$ $\qquad 2(1 - pt(2.174, 29)) = 0.038$

2 d) We reject the null hypothesis at the C=95% confidence level, i.e. test shows there is significant evidence average is different for NYers.

5. In the previous example, find the power of the test to distinguish the New Yorkers from the general population, if the actual number of hot dogs New Yorkers eat each month really is 8.2.

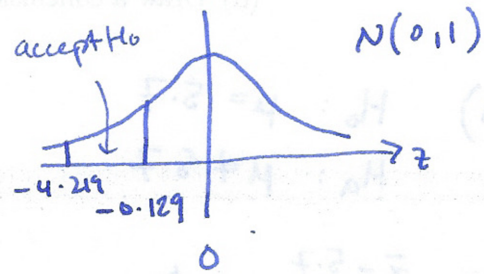
critical values for test are $5.7 \pm t_{29}^{7} \times \frac{6.3}{\sqrt{30}}$ $\quad t_{29}^{7} = 2.045$

$3.348,\ 8.052$



convert to standard normal

$z = \frac{w - 8.2}{6.3/\sqrt{30}}$



power = prob reject Ho given $\mu = 8.2$ = $1 - \text{pnorm}(-0.129) + \text{pnorm}(-4.219)$

$1 - 0.0985 + 0$

$= 90\%.$

↑ should really use t_{29} but don't have enough values in table so approx by normal here.

6. You wish to take a sample of CSI students to find what proportion of them use iPhone's. How large does your sample need to be to have a margin of error less than 10%? You may use the fact that $p(1-p) \leq \frac{1}{2}$.

assume $C = 95\%$

margin of error $z^* \sqrt{\frac{p(1-p)}{n}} \leq z^* \sqrt{\frac{1/2}{n}}$ $z^* = 1.96$

$1.96 \sqrt{\frac{1/2}{n}} = 0.1$

$\frac{(1.96)^2}{2n} = (0.1)^2$

$n = \frac{(1.96)^2}{2(0.1)^2} = 192$

7. You take a survey of 120 CSI students and find that 78 of them have iPhone's. Find a 95% confidence interval for the proportion of students who have iPhone's.

$\hat{p} = \frac{78}{120} = 0.65$

confidence interval: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$0.\cancel{0}65 \pm 1.96 \sqrt{\frac{0.65 \times 0.35}{120}} = (0.56, 0.74)$

8. A national survey of 1000 people finds that 523 of them have an iPhone. Find a 95% confidence interval for the difference between the national proportion, and the proportion of CSI students using the sample in the previous question.

$\hat{p}_1 = 0.65$ $\hat{p}_2 = \frac{523}{1000} = 0.523$ $\hat{p} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} = \frac{78 + 523}{1120} = 0.5366$

confidence interval $\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$0.127 \pm 1.96\sqrt{0.5366 \times (1 - 0.5366)\left(\frac{1}{120} + \frac{1}{1000}\right)}$

$(0.033, 0.221)$

9. Using the data from the previous two questions, test whether CSI students are significantly more likely to use an iPhone compared the national population. (Choose a C = 98% confidence level.) State your null and alternative hypotheses explicitly, and also compute the p-value.

$H_0: p_1 - p_2 = p = 0$

$H_a:$ ~~$p_1 - p_2 \neq 0$~~ $p_1 > p_2,\ p > 0$

test statistic $\sim N(0,1)$ $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \dfrac{0.127}{\sqrt{0.5366(1-0.5366)\left(\frac{1}{120}+\frac{1}{1000}\right)}} = 2.636$

p-value: $1 - \text{pnorm}(2.636) = 0.0042$

significant at $C = 98\%$ confidence level.

conclusion: reject H_0

i.e. significant evidence CSI students more likely to use iPhone.

10. It is known that the overall proportion of cell phones in the US which are iPhone's is 52.28%. If you choose a sample of CSI students of size 120, what is the power of the test in the previous question to detect an alternative hypothesis H_a: $p = 78/120$?

z^* for 1-sided 98% test

test rejects H_0 if test statistic ≥ 2.06

if $p_1 = 0.65$ then $\hat{p}_1 \sim N\left(0.65, \sqrt{\frac{0.65(1-0.65)}{120}}\right)$

$$\text{prob}\ \frac{\hat{p}_1 - 0.5228}{\sqrt{\hat{p}_1(1-\hat{p}_1)/120}} \geq 2.06 = \text{prob}\ \hat{p}_1 \geq 2.06\sqrt{\hat{p}_1(1-\hat{p}_1)/120} + 0.5228$$

$$\geq 0.612$$

convert to $N(0,1)$: $\quad$ prob $N(0,1) \geq \frac{0.612 - \cancel{0.5228}\ \cancel{0.65}\ 0.65}{\sqrt{\frac{0.65(1-0.65)}{120}}} = \cancel{-0.047}\ \cancel{-0.047}\ -0.873$

power $= 1 - \text{pnorm}(\cancel{-0.047}\ -0.873) = \cancel{82\%}.\ 0.81 \quad 81\%$

Formulas

μ population mean
σ population standard deviation
n sample size
$\overline{x}$ sample mean
s sample standard deviation (standard error)
p population proportion
$\hat{p}$ sample proportion

The sample mean $\overline{x}$ of a normal distribution $N(\mu, \sigma)$ has distribution $N(\mu, \sigma/\sqrt{n})$.

The sample mean of any distribution with mean μ and standard deviation σ has distribution approximately $N(\mu, \sigma/\sqrt{n})$, for n sufficiently large.

	Confidence interval	Test statistic	Distribution
mean, known σ	$\overline{x} \pm z_* \sigma/\sqrt{n}$	$\dfrac{\overline{x}-\mu}{\sigma/\sqrt{n}}$	$N(0,1)$
mean, unknown σ	$\overline{x} \pm t_* s/\sqrt{n}$	$\dfrac{\overline{x}-\mu}{s/\sqrt{n}}$	t-dist, df $= n-1$
difference between two means	$\overline{x_1} - \overline{x_2} \pm t_* \sqrt{s_1^2/n_1 + s_2^2/n_2}$	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$	t-dist, df=$\min\{n_1, n_2\} - 1$
proportion	$\hat{p} \pm z_* \sqrt{\hat{p}(1-\hat{p})/n}$	$\dfrac{\hat{p}-p}{\sqrt{p(1-p)/n}}$	$N(0,1)$
difference between two proportions	$\hat{p}_1 - \hat{p}_2 \pm z_* \sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$, where $\hat{p} = (\hat{p}_1 n_1 + \hat{p}_2 n_2)/(n_1+n_2)$	$\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$	$N(0,1)$

$$\chi^2 \text{ statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$