Math 214 Statistics Spring 19 Final

Name: Solutions

- You may use a 3 × 5 index card of notes, and a calculator.
- 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	

Final	
Overall	

1. (a) Suppose you toss a fair coin three times. What is the probability that you get an even number of heads?

   (b) In the US population, the proportion of left-handed people is 9.8%, the proportion of red-haired people is 2.2%, and the proportion of the population which is male is 49.2%. If all these characteristics are independent, find the probability someone you pick at random is female with red hair and right-handed.

a) HHH
HHT —
HTH —
HTT
THH —
THT
TTH —
TTT

$\frac{4}{8} = \frac{1}{2}$

b) $(1-0.492)(0.022)(1-0.098) = 1.008075\%$.

2. Find the mean, median, standard deviation, Q_1 and Q_3 for the following set of numbers: $1, 1, 4, 5, 8, 10, \cancel{19}.$ 20

$$\text{mean} = \frac{1+1+4+5+8+10+\overset{20}{\cancel{19}}}{7} = \frac{49}{7} = \cancel{6.857}\ 7$$

$$\text{median} = 5$$

$$\text{s.d.} = \sqrt{(6^2+6^2+3^2+2^2+1^2+3^2+13^2)/7} = 6.071$$

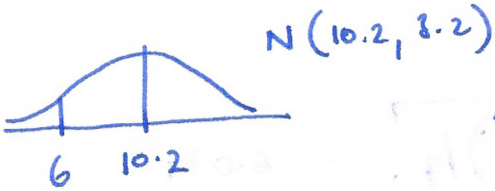
$$Q_1 = 1$$

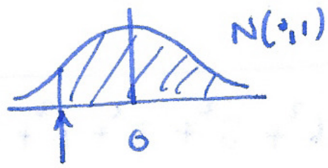
$$Q_3 = 10$$

3. The height of trees in a particular forest is normally distributed with mean 10.2 m and standard deviation 3.2 m.

   (a) If you pick a tree at random, what is the probability that it is taller than 6 m?

   (b) How large does a random sample need to be for a 95%-confidence interval for the mean to have a margin of error of at most 1 m?

a) 

$\rightsquigarrow$ 

$$\frac{6-10.2}{3.2} = -1.3125$$

$$1 - \text{pnorm}(-1.3125) = 1 - 0.0951 = 0.9059$$

b) confidence interval: $\bar{x} \pm z^* \sigma/\sqrt{n}$

want $\dfrac{1.96 \cdot 3.2}{\sqrt{n}} = 1$ $\qquad n = ((1.96)\times(3.2))^2 = 40$

4. You take a simple random sample of 21 radiology technicians in NYC and discover their mean income is \$76,020, with standard deviation \$12,756. Find a 95% confidence interval for the actual mean income of radiology technicians in NYC.

confidence interval $\bar{x} \pm t^* s/\sqrt{n}$ t^* for $df = 20$ is $t^* = 2.086$

$76{,}020 \pm 2.086 \times 12{,}756/\sqrt{21} = (70{,}213, 81{,}827)$

5. You take a random sample of size 500 from a population with unknown mean μ and unknown standard deviation σ. The sample mean is $\overline{x}$, and the standard deviation of the sample is s. You run a hypothesis test with $H_0 : \mu = 10$ and $H_a : \mu \neq 10$, and the resulting p-value is 0.26.

Indicate whether or not the following statements are true or false.

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

○ True
○ False

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

○ True
○ False

(c) The standard deviation s of the sample should be close to $\sigma/\sqrt{500}$.

○ True
○ False

(d) The probability that the null hypothesis is false is 0.26.

○ True
○ False

(e) This is evidence that the population mean μ is close to 10.

○ True
○ False

6. A random sample of 14 cars in NY average 27.2 mpg, with standard deviation 5.3 mpg, while a random sample of 17 cars in NJ average 23.2 mpg, with standard deviation 4.2 mpg. Find a 95%-confidence interval for the difference between the fuel efficiency of cars in NY and NJ.

t_α for $df = 13$ is $t_\alpha = 2.160$

$$\bar{x}_1 - \bar{x}_2 \pm t_\alpha \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

$$27.2 - 23.2 \pm 2.160 \times \sqrt{5.3^2/14 + 4.2^2/17}$$

$$= (0.231, 7.769)$$

7. You take a random sample of 34 CSI students, and 25 claim to be following the current season of Game of Thrones. A random sample of 43 New Yorkers indicates that 27 of them are following the current season of Game of Thrones. Set up and implement a 95%-significance level hypothesis test to see if this is evidence that different numbers of CSI students watch Game of Thrones as compared to the overall NYC population.

$\hat{p}_1 = \frac{25}{34} = 0.735 \qquad \hat{p}_2 = \frac{27}{43} = 0.628$

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

test statistic: $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $\hat{p} = \frac{25+27}{34+43} = 0.675$

$= \dfrac{0.735 - 0.628}{\sqrt{0.675(1-0.675)\left(\frac{1}{34}+\frac{1}{43}\right)}} = 0.995$ compare w/ $z^* = 1.96$

no significant difference at 95%-confidence level.

8. A study investigates the relation between salt intake and high blood pressure. A random sample of the population is taken, and they are classified as low salt intake or high salt intake depending on their diet, and as to whether or not they have high blood pressure.

	Low salt	High salt
High blood pressure	9	19
Not high blood pressure	103	92

Here is the output from R:

```
> d <- matrix( c(9, 19, 103, 92), ncol=2, nrow=2  )
> chisq.test(d)

Pearson's Chi-squared test with Yates' continuity correction

data:  d
X-squared = 3.4013, df = 1, p-value = 0.06515
```

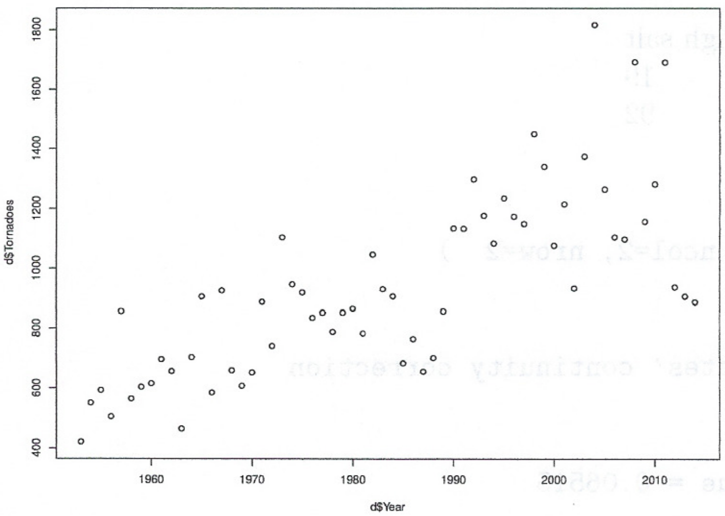
(a) What do you conclude from the R output?

(b) Would you recommend the researchers repeat the study with a larger sample size? Explain.

a) There's no significant difference at the 95% confidence level.

b) The p-value is close to 0.05 the standard significance level, might be worth repeating with a large sample size.

9. You obtain data on the number of tornadoes reported each year from 1953 to 2014. Here is a plot of your data and the R output:




```
> summary(lm(d$Year~d$Tornadoes))
Call:
lm(formula = d$Year ~ d$Tornadoes)
Residuals:
     Min       1Q   Median       3Q      Max
-22.744   -8.584   -1.356    5.048   32.827
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.942e+03  4.915e+00 395.029  < 2e-16 ***
d$Tornadoes 4.466e-02  4.975e-03   8.975 1.09e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 11.89 on 60 degrees of freedom
Multiple R-squared:  0.5731,Adjusted R-squared:  0.566
F-statistic: 80.56 on 1 and 60 DF,  p-value: 1.086e-12
```

(a) What is the best fit regression line?

(b) Does this give evidence that the number of tornadoes is increasing? Can you think of any other reasons why the number of observed tornadoes might be increasing?

a) $T = 1942 + 0.0466Y$.

b) The data shows strong evidence for a small increase with year, but this might just be due to the fact that the population is increasing and so more tornados are observed.

10. Assume the distribution of women's shoe sizes in the US is normally distributed with mean 7.9 and standard deviation 1.7.

Suppose you take a simple random sample of 36 female CSI students. Describe a test with a 95% confidence level to decide whether or not they have same average shoe size as the general population.

What is the power of your test to detect the fact that the actual shoe size of female CSI students is normally distributed with mean 8.1 and standard deviation 1.7?

$H_0 : \mu = 7.9$

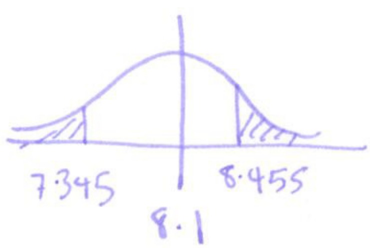
$H_a : \mu \neq 7.9$

test statistic : $\dfrac{\bar{x} - 7.9}{1.7/6} \sim N(0,1)$

compare w/ ± 1.96.

critical values for $\bar{x}$: $7.9 \pm 1.96 \times 1.7/6 = (7.345, 8.455)$

power :



$= \text{pnorm}\left(\dfrac{7.345 - 8.1}{1.7/6}\right) + 1 - \text{pnorm}\left(\dfrac{8.455 - 8.1}{1.7/6}\right)$

$\cancel{-0.281}$ $\cancel{0.209}$

-2.665 1.253.

$= \cancel{7.65\%}$

10.9%.