

Stat 254 - Final Exam Review

1. A hypothesis test of $H_o : \mu = 25$ vs $H_a : \mu > 25$ yielded a p-value of 0.357. Explain in plain language what this result means.
2. 22.3% of Canadians speak French as their first language. In a random sample of 12 Canadians, we want to determine the number that speak French as their first language.
 - (a) Define the random variable, state its distribution, and find the mean and variance of the random variable.
 - (b) What is the probability that exactly 2 in the sample speak French as their first language?
 - (c) What is the probability that at least 5 in the sample speak French as their first language?
3. What is the difference between a statistic and a parameter? Give an example of each.
4. The amount of time students spend studying for their introductory statistics class each week is normally distributed with mean 3.5 hours and a standard deviation of 0.45 hours. Select each answer from the given list.
 - (a) What is the probability that a randomly selected student studies less than 2.25 hours each week?
(a) 0.0010 (b) 0.0027 (c) 0.0523 (d) 0.2583
 - (b) What is the probability that a randomly selected student studies more than 3 hours each week?
(a) 0.1032 (b) 0.4888 (c) 0.6033 (d) 0.8665
 - (c) One student claims that she is always studying and is in the top 2% of students in terms of study time. What is the minimum number of hours she must study each week?
(a) 2.23 (b) 3.5 (c) 4.42 (d) 5.0
 - (d) Take a random sample of 20 students. What is the probability that their average study time is within 0.25 hours of the mean?
(a) 0.4218 (b) 0.7323 (c) 0.9870 (d) 0.9932
 - (e) Suppose that 15.87% of students that study less than 3 hours per week. Take a random sample of 40 students, what is the probability that at least 20% of students in the sample study less than 3 hours per week?
(a) 0.0211 (b) 0.1010 (c) 0.2374 (d) 0.7627
5. We wish to survey Camosun students to see if they are happy with the current student government. If a 99% confidence level and 0.03 maximum error are desired, determine the required sample size. Assume that we have no idea what the actual proportion might be. Choose the critical value from the given list.

<pre>> pnorm(0.995,0,1)</pre>	<pre>> qnorm(0.995,0,1)</pre>
<pre>[1] 0.8401319</pre>	<pre>[1] 2.575829</pre>
<pre>> qnorm(0.99,0,1)</pre>	<pre>> qt(0.995,98)</pre>
<pre>[1] 2.326348</pre>	<pre>[1] 2.626931</pre>
6. What happens to the width of a confidence interval when you (a) increase confidence level? (b) increase sample size?
7. To test a new style of running shoe for comfort, researchers asked 20 people who used the new style if they found the shoes to be comfortable and also asked 20 people who used other shoes and then compared the reported comfort level. Those wearing the new shoes reported higher comfort levels than those who used other shoes. Explain why this does not prove that the new shoes are more comfortable than other shoes.

8. For the following data

9 16 12 16 15 23 19 17 35

- (a) Are there any outliers?
 - (b) Draw a boxplot and comment on its shape.
9. What is the difference between a point estimate and an interval estimate? Give an example of each.
10. A company is willing to renew its advertising contract with a local radio station only if the station can prove that more than 20% of the residents of the city have heard the ad and recognize the company's product. The radio station conducts a random phone survey of 400 people.
- (a) What are the hypotheses?
 - (b) The station plans to conduct this test using a 10% level of significance, but the company wants the significance level lowered to 5%. Why?
 - (c) What is meant by the power of this test?
 - (d) For which of the two levels of significance will the power of this test be higher? Why?
11. The research division of a company claims that a new pain relief medicine they have developed provides faster relief than the current standard mean of 39 minutes. The company will begin to manufacture the new medicine if there is strong evidence that this claim is true. Choose your p-value and critical values from the list below.
- (a) State the null and alternative hypotheses the company would use to test this claim.
 - (b) What are the consequences of type I and type II errors the company could make.
 - (c) The new medicine is tested on a random sample of 36 individuals resulting in a mean time to relief of 34 minutes and standard deviation 13 minutes. Is there enough evidence to conclude that the average time to relief is less than 39 minutes for the new medicine at $\alpha = 0.05$? Support your answer with a p-value and a 95% confidence interval.

```
> qt(0.975,35)
[1] 2.030108
> pt(abs(statistic),df)
[1] 0.9864855
> pnorm(abs(statistic))
[1] 0.9895004
> qnorm(0.975,0,1)
[1] 1.959964
> qt(0.95,35)
[1] 1.689572
> qt(0.975,36)
[1] 2.028094
```

12. A Carrier XH heat pump, designed to heat a home in the winter and cool the home in the summer, lasts, on average, 16 years. The lifetime (in years) can be modelled by an exponential random variable. What is the probability that a randomly selected heat pump will last for at least 5 years?

13. A survey suggests that 65% of Canadians feel stress from work or school. A survey of 586 Camosun students found that 395 of them felt stressed. Is the proportion of stressed out students at Camosun higher than the national proportion? Test the claim at the $\alpha = 0.05$ significance level and calculate a 95% confidence interval to estimate the true proportion of stressed students. (Define the parameter, State your hypothesis, check any assumptions, calculate the statistic and p-value, find a 95% confidence interval, and state your conclusions based on both the p-value and the confidence interval.)

```

> qt(0.975,585)
[1] 1.964027
> pnorm(abs(statistic))
[1] 0.8887676
> pt(abs(statistic),df)
[1] 0.8981122
> qnorm(0.95,0,1)
[1] 1.644854
> qnorm(0.975,0,1)
[1] 1.959964
> qchisq(0.975,585)
[1] 653.9148

```

14. Recent evidence suggests that nonsmokers who live with smokers and are therefore exposed to second-hand smoke (as opposed to first-hand smoke that is directly inhaled by the smoker) may experience adverse health effects. The following are data on tar yield for 8 brands of Canadian cigarettes. (a) Is there enough evidence at $\alpha = 0.05$ to conclude that the mean tar yield from second hand smoke is different from that of first hand smoke? (b) Find a 95% confidence interval to estimate the true mean difference in tar yield. (Define the parameter, State your hypothesis, check any assumptions, calculate the statistic and p-value, find a 95% confidence interval, and state your conclusions based on both the p-value and the confidence interval.)

Cigarette Brand	1	2	3	4	5	6	7	8
2nd Hand Yield	15.8	16.9	21.6	18.8	29.3	20.7	18.9	25.0
1st Hand Yield	18.5	17.0	17.2	19.4	15.6	16.4	13.3	10.2

```

> qt(0.975,8)
[1] 2.306004
> qt(0.975,7)
[1] 2.364624
> qnorm(0.975,0,1)
[1] 1.959964
> pnorm(0.95,0,1)
[1] 0.8289439
> pnorm(abs(statistic))
[1] 0.9848444
> pt(abs(statistic),df)
[1] 0.9664942

```

15. The random variables X has the distribution

x	0	3	5	8
$p(x)$	0.2	0.25	0.45	.1

- (a) Is this a legitimate probability distribution? Explain.
 (b) What is the probability of that X is more than 3?
 (c) Find the mean, variance, and standard deviation of X .

16. According to Nielsen media research, children (ages 2-11) spend an average of 21.5 hours watching tv per week, while teens (ages 12-17) spend an average of 20.7 hours. Based on the sample statistics obtained below, is there sufficient evidence to conclude a difference in average tv watching times at $\alpha = 0.01$? (Define the parameter, State your hypothesis, check any assumptions, calculate the statistic and p-value, find a 99% confidence interval, and state your conclusions based on both the p-value and the confidence interval.)

Teens	$n_1 = 15$	$\bar{x}_1 = 18.50$	$s_1 = 1.10$
Children	$n_2 = 15$	$\bar{x}_2 = 22.45$	$s_2 = 4.05$

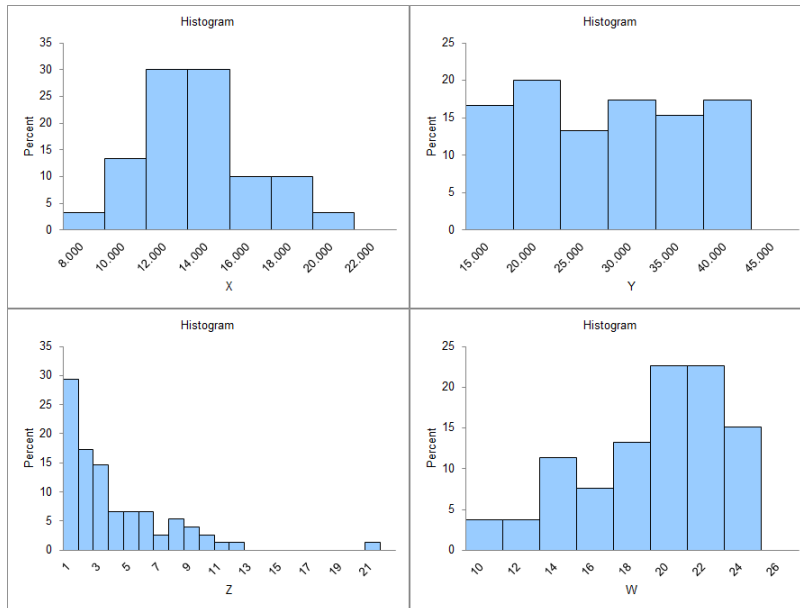
```

> qnorm(0.995,0,1)
[1] 2.575829
> qnorm(0.99,0,1)
[1] 2.326348
> pnorm(abs(statistic))
[1] 0.9998663
> pt(abs(statistic),df)
[1] 0.9986745
> qt(0.995,29)
[1] 2.756386
> qt(0.99,14)
[1] 2.624494
> qt(0.995,14)
[1] 2.976843

```

17. Let A be the event that a kitten is striped and B be the event that a kitten is female. Assume $P(A) = 0.3$ and $P(B) = 0.5$ and $P(A \text{ and } B) = 0.15$
- Describe what it means to be mutually exclusive in plain language
 - In terms of probability, what must be true if A and B are mutually exclusive?
 - Describe what it means to be independent in plain language
 - In terms of probability, what must be true if A and B are independent?
 - Find the probability of a kitten that is striped or female.
 - Find the probability of a kitten that is female with stripes.
 - Find the probability of a kitten being female, given that it is striped.
 - Find the probability of a kitten being striped, given that she is female.
18. A pet store sells both food and other supplies such as pet beds. Suppose 45% of customers at this store own cats and 55% own dogs. The store owners have observed 68% of the cat owners buy food when they visit the store, and the rest buy other supplies; 57% of dog owners buy food, and the rest buy other supplies.
- Draw a tree diagram to represent this problem.
 - What is the probability that a randomly selected customer is a dog owner that buys other supplies?
 - What is the probability that a randomly selected customer buys other supplies?
 - If a randomly selected customer buys food, then what is the probability that customer is a cat owner?

19. Describe the distributions exhibited by each of the following histograms



20. A criminology student wants to determine if a relationship exists that so we can use robbery rates to predict homicide rates in Canada. Data was collected from 10 randomly selected cities, reported in crime rate per 100,000 population and has data output as shown below.

City	1	2	3	4	5	6	7	8	9	10
Robbery Rate	229	148	141	125	91	88	84	173	67	152
Homicide Rate	4.9	2.6	3.4	1.7	1.9	1.3	1.1	3.7	0.7	3.1

```
> summary(modelA)
```

Call:

```
lm(formula = homicide ~ robbery)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-0.61558 -0.14265 -0.08394  0.12629  0.66968
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.924559   0.363541  -2.543   0.0345 *
robbery      0.025921   0.002634   9.840 9.57e-06 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3905 on 8 degrees of freedom
Multiple R-squared:  0.9237,    Adjusted R-squared:  0.9141
F-statistic: 96.83 on 1 and 8 DF,  p-value: 9.573e-06
```

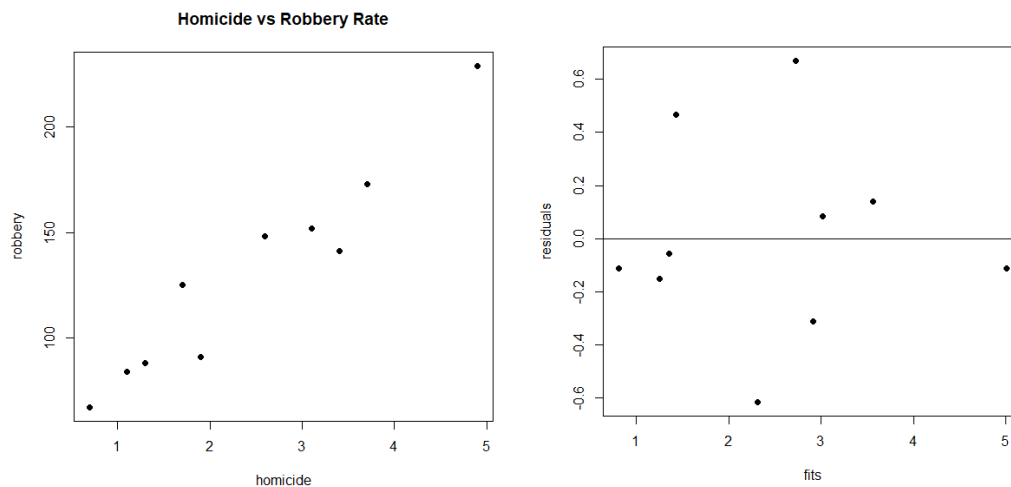
```
> summary(modelB)

Call:
lm(formula = robbery ~ homicide)

Residuals:
Min      1Q  Median      3Q      Max
-23.0090  -1.6042  -0.9865   9.1421  21.5694

Coefficients:
(Intercept)  42.852      9.952   4.306  0.00259 **
homicide      35.634      3.621   9.840 9.57e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.48 on 8 degrees of freedom
Multiple R-squared:  0.9237,    Adjusted R-squared:  0.9141 
F-statistic: 96.83 on 1 and 8 DF,  p-value: 9.573e-06
```



- (a) From the R output above, which model should be used for predicting homicide - modelA or modelB?
 - (b) Calculate and interpret the correlation coefficient.
 - (c) Calculate the equation of the linear regression line.
 - (d) What change in homicide rate is associated with an increase by 1 in the robbery rate?
 - (e) Find and interpret the coefficient of determination.
 - (f) Estimate the predicted homicide rate when robbery rate is 141 and find the residual for this observation.
21. According to recent FBI statistics, the mean number of bank robberies per day in the southern region of the United States is 4.32. Suppose a day is selected at random, what is the probability that
- (a) exactly 2 bank robberies occur in the southern region on that day?
 - (b) at least 2 bank robberies occur in the southern region on that day?

22. According to Major League Baseball rules, a baseball should weight between 5 and 5.25 ounces. Suppose the weight of a baseball has a uniform distribution between 5.085 and 5.155 ounces. A baseball is selected at random.

- (a) What is the probability that baseball weighs more than 5.100 ounces?
- (b) What is the probability that baseball weighs between 5.090 and 5.100 ounces?
- (c) What is the cutoff weight for a baseball to be in the lowest 10% in terms of weight?

23. At a sheep farm, we sheared 53 Columbian lambs and weighed their wool. The sample yielded an average fleece weight of 12 lbs with a standard deviation of 4 lbs. The farmer claims that his lambs have an average fleece of 14 lbs. Conduct a test to determine if the weight is actually lower than the farmer's claim at $\alpha = 0.01$. (State your hypothesis, check any assumptions, calculate the statistic and p-value, find a 99% confidence interval, and state your conclusions based on both the p-value and the confidence interval.)

```

> pnorm(abs(statistic))
[1] 0.9998637
> pt(abs(statistic),df)
[1] 0.9996861
> qnorm(0.995)
[1] 2.575829
> qt(0.995,52)
[1] 2.673734
> qt(0.99,53)
[1] 2.39879

```

24. Some people seem to believe that you can fix anything with duct tape. Even so, many were skeptical when researchers announced that duct tape may be a more effective and less painful alternative than liquid nitrogen, which doctors routinely use to freeze warts. A study of the effectiveness of the procedure random assigned patients with warts to either have the duct tape treatment or the more traditional freezing treatment. Of the 100 that had the freezing treatment, 60 successfully had the wart removed. Of the 104 that used the duct tape method, 83 had the wart successfully removed. Is there significant statistical evidence that the freezing is less effective than the duct tape treatment? Test the claim at $\alpha = 0.01$ and construct a 99% confidence interval to estimate the difference. (State your hypothesis, check the conditions, calculate the statistic and p-value, find a 99% confidence interval, and state your conclusions based on both the p-value and the confidence interval.)

```

> pnorm(abs(statistic))
[1] 0.9989962
> pt(abs(statistic),df)
[1] 0.998697

```