1. If the average really was 25, like we thought, the data we got would only occur about 36% of the time. That is more than 1 in every 3 times, which is pretty common. Since our data were so close to what we expected to see, that means we have not really found much reason to doubt the claim and we will continue assuming it is true.

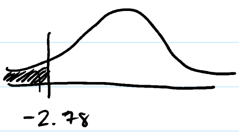
2. a) X = number with French as first language out of 12 Canadians
$X \sim \text{Binom}(12, 0.223)$
$\mu = np = 12(.223) = 2.676$ $\quad \sigma^2 = np(1-p) = 12(.223)(.777) = 2.08$

b) $P(X=2) = \binom{12}{2} .228^2 .777^{10} = 0.2632$

c) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) - P(X=4)$
$= 1 - \binom{12}{0} .228^0 .777^{12} - \binom{12}{1} .228^1 .777^{11} - \ldots - \binom{12}{4} .228^4 .777^8$
$= 1 - .0484 - .1668 - .2632 - .2518 - .1777$
$= 0.1071$

3. Statistic summarizes a sample $\rightarrow \bar{x}, s, \hat{p}$
Parameter summarizes a population $\rightarrow \mu, \sigma, p$

4. a) $P(X < 2.25) = P\left(Z < \frac{2.25 - 3.5}{0.45}\right) = P(Z < -2.778)$



z in $(-3, -2)$
so prob. in $(0.0015, 0.025)$ $\rightarrow$ $\boxed{0.0027}$

b) $P(X > 3) = P\left(Z > \frac{3 - 3.5}{0.45}\right) = P(Z > -1.11)$

Z in $(-2, -1)$ $\Rightarrow$ prob in $(84\%, 97.5\%)$ $\rightarrow$ $\boxed{0.8665}$

c) $P(X > a) = 0.02$ $\rightarrow$ needs z between $+2$ and $+3$

so X between $3.5 + 2(.45) = 4.4$
and $3.5 + 3(.45) = 4.85$

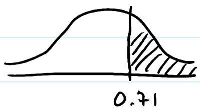
$\boxed{a = 4.42}$

d) $P(3.25 < \bar{x} < 3.75) = P\left(\frac{3.25 - 3.5}{0.45/\sqrt{20}} < Z < \frac{3.75 - 3.5}{0.45/\sqrt{20}}\right) = P(-2.48 < Z < +2.48)$

Between 95% and 99.7% $\rightarrow$ $\boxed{0.9870}$

e) $P(\hat{p} > 0.2) = P\left(Z > \frac{0.2 - 0.1587}{\sqrt{\frac{0.1587(0.8413)}{40}}}\right) = P(Z > 0.71)$

Since $np = 40(.1587) = 6.335$
$n(1-p) = 40(.8413) = 33.7$



Z in $(0, 1)$ so prob. in $(.16, .5)$
$\rightarrow$ $\boxed{0.2374}$

5. $n = \hat{p}(1-\hat{p})\left(\frac{z^*}{B}\right)^2 = 0.5(0.5)\left(\frac{2.5758}{0.03}\right)^2 = 1842.98$

use n at least 1843.

6. a) Wider b) narrower

7. - Did not specify the type/age/quality of shoe used by the "control" group.
- Did not fix the activities done by either group
- Many possible confounding variables

8. a) 9 12 15 16 16 17 19 23 35

Quartiles: $d_1 = 9/4 = 2.25 \rightarrow$ position 3 $\rightarrow Q_1 = 15$

$d_3 = 9 \times 3/4 = 6.75 \rightarrow$ position 7 $\rightarrow Q_3 = 19$

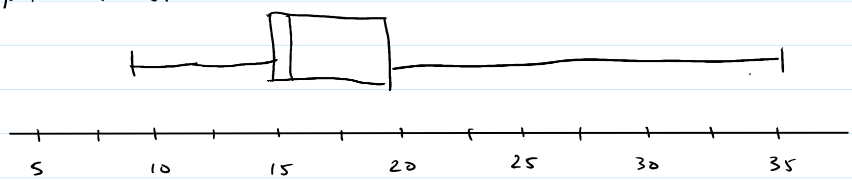
IQR = 19 − 15 = 4

lower = 15 − 1.5(4) = 9

upper = 19 + 1.5( ) = 25

The value 35 is a potential outlier since it is outside the fences

b) Median = 15



Skewed right

9. A point estimate is a single value from a sample used to estimate the parameter (eg $\bar{x}$) while an interval estimate gives a collection of potential values (eg $\bar{x} \pm t^* s/\sqrt{n}$)

10. a) $H_0: p = 0.2$ $H_a: p > 0.2$

b) Rejecting Ho means more than 20% of residents hear the ad and it should be worth implementing. Since it is beneficial to the station to sell the ad space, they would want a higher chance of rejecting Ho (higher α). For the company, they don't want to risk buying ads that are not reaching the audience so they would want a lower value of α.

c) The probability that the sample has significantly more than 20% of residents reached when the ad is truly effective

d) As α increases, power also increases

11. a) $H_0: \mu = 39$ $H_a: \mu < 39$

b) Type I − Reject Ho when Ho is true
- They would manufacture the new medicine but it does not actually improve relief time

Type II − fail to reject Ho when Ho is false.
- They would chose not to manufacture the new medicine but it would have actually been an improvement.

c) $t_0 = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{34 - 39}{13/\sqrt{36}} = -2.308$ $\nu = 36 - 1 = 35$

p-value = 1 − 0.9865 = 0.0135 (using pt(2.308, 35)

There is significant evidence to suggest the new medicine has mean relief time lower than 39 minutes and they should

There is significant evidence to suggest the new medicine has mean relief time lower than 39 minutes and they should start producing it.

$$\bar{x} \pm t^* s/\sqrt{n} = 34 \pm 2.0301 \left(13/\sqrt{36}\right) = 34 \pm 4.399 = (29.6, 38.4)$$

We are 95% confident the true mean relief time of the new medicine is 29.6 to 38.4 minutes. Since this interval is entirely below 39, it agrees with our conclusion from the test.

12. $$\int_5^\infty \frac{1}{16} e^{-\frac{1}{16}x} dx = -e^{-\frac{1}{16}x}\Big|_{x=5}^{\infty} = 0 - \left(-e^{-5/16}\right) = 0.7316$$

13. $H_0: p = 0.65$ $\quad H_a: p > 0.65$

for p = proportion of Camosun students that are stressed.

$n\hat{p} = 395 \quad n(1-\hat{p}) = 191 \quad > 5$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{395/586 - 0.65}{\sqrt{\frac{.65(.35)}{586}}} = 1.22$$

p-value = 0.1112

There is not enough evidence to reject H_0. We may assume that 65% of Camosun students are stressed, no higher than the national proportion.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 395/586 \pm 1.96 \sqrt{\frac{395/586\,(191/586)}{586}}$$

$$= 0.674 \pm 0.038 = (0.636, 0.712)$$

We are 95% confident that 63.6% to 71.2% of all Camosun students are stressed. Since this interval is not strictly above 65%, it agrees with our conclusion from the hypothesis test

14. Differences: $d = 1^{st} - 2^{nd} = \{2.7, 0.1, -4.4, 0.6, -13.7, -4.3, -5.6, -14.8\}$

$\bar{d} = -4.925$ $S_d = 6.432$ $n = 8$

$H_0: \mu_d = 0$ $H_a: \mu_d \neq 0$

Conditions: Assume Normally distributed (check Normal Probability plot)

Statistic: $t_0 = \dfrac{\bar{d} - \mu_d}{S_d/\sqrt{n}} = \dfrac{-4.925 - 0}{6.432/\sqrt{8}} = -2.166$ $df = 7$

p-value = 0.0670 using 2xPt(-2.166, 7)

Conclusion: p-value > $\alpha = 0.05$. There is insignificant evidence to suggest that there is a difference, on average, between 1^{st} and 2^{nd} hand smoke tar yield.

Interval: $\bar{d} \pm t_{\alpha/2, n-1} S/\sqrt{n} = -4.925 \pm 2.3646\left(\frac{6.432}{\sqrt{8}}\right) = -4.925 \pm 5.377$

$= (-10.302, 0.452)$

We are 95% confident that the mean difference in tar yield for 1^{st} and 2^{nd} hand smoke is between -10.302 and 0.452.

Since the interval includes 0, there is not enough evidence to suggest there is a difference, on average.

15 a) yes: $0 \leq p_i \leq 1$ for all i, $\sum p_i = 1$

b) $P(X \geq 3) = 0.45 + 0.1 = 0.55$

c) $\mu = \sum x \cdot p = .2(0) + .25(3) + .45(5) + .1(8) = 3.8$

$\sigma^2 = \sum (x-\mu)^2 \cdot p = .2(0-3.8)^2 + .25(3-3.8)^2 + .45(5-3.8)^2 + .1(8-3.8)^2$

$\sigma^2 = 5.46$ $\sigma = \sqrt{\sigma^2} = 2.34$

16 μ_c = mean TV time for children μ_T = mean TV time for teens

$H_0: \mu_c - \mu_T = 0$ $H_a: \mu_c - \mu_T \neq 0$

Conditions: Check Norm. prob. plot since $n < 30$, σ unknown.

Statistic: $t_0 = \dfrac{(\bar{X}_c - \bar{X}_T) - \Delta_0}{\sqrt{S_c^2/n_c + S_T^2/n_T}} = \dfrac{(22.45 - 18.50) - 0}{\sqrt{4.05^2/15 + 1.10^2/15}} = 3.645$

(cont'd) $t_0 = 3.645$ $df = \min(15-1, 15-1) = 14$

P-value = 0.0046 $2 * pt(-3.645, 14)$

Conclusion: p-value $< \alpha = 0.01$. There is very strong evidence to suggest that there is a difference in the true mean TV time for teens and children.

Interval: $(\bar{X}_c - \bar{X}_T) \pm t_{\alpha/2, df}\sqrt{(s_c^2/n_c + s_T^2/n_T)} = (22.45 - 18.5) \pm 2.9768\sqrt{\frac{4.05^2}{15} + \frac{1.10^2}{15}}$

$= 3.95 \pm 3.23 = (0.72, 7.18)$

We are 99% confident that the true mean TV time for children is between 0.72 and 7.18 hours higher than for teens.

17. a) It is impossible for both events to occur.

b) $P(A \cap B) = 0$ so $P(A \cup B) = P(A) + P(B)$

c) Knowing that A occurs does not change the probability that B occurs.

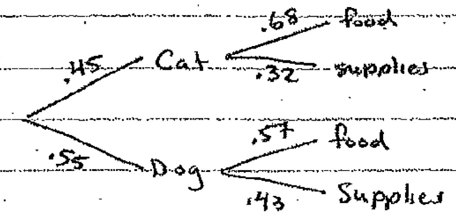
d) $P(A \cap B) = P(A) \times P(B)$ also $P(A|B) = P(A)$

e) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .3 + .5 - .15 = .65$

f) $P(A \cap B) = 0.15$

g) $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{.15}{.3} = .5$ h) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.15}{.5} = 0.3$

18 a)



b) $P(D \cap S) = .55 \times .43 = .2365$

c) $P(S) = .32 \times .45 + .43 \times .55 = 0.3805$

d) $P(C|F) = \frac{P(C \cap F)}{P(F)} = \frac{.68 \times .45}{.68 \times .45 + .57 \times .55} = 0.4939$

19 top left – Symmetric, unimodal, peak at 14000, no outliers

top right – flat (uniform), no outliers

bottom left – skewed right, unimodal – peak at 1, outlier at 21

bottom right – skewed left, unimodal – peak at 22, no outliers.

20 a) model A $x =$ robbery

b) $R^2 = 0.9237$ so $R = \pm\sqrt{.9237} = 0.961$

The linear association is positive and strong.

c) $\hat{y} =$

b) $R^2 = 0.9237$ so $R = \sqrt{0.9237} = 0.961$

The linear association is positive and strong.

c) $\hat{y} = -0.92 + 0.026x$

d) $b_1 = 0.026$

e) $R^2 = 0.9237 \rightarrow$ About 92% of the variability in homicide rate is explained using the linear model with robbery rate

f) $\hat{y} = -0.92 + 0.026(141) = 2.746$ (obs. 3)

$y = 3.4$ so Residual: $e = y - \hat{y} = 3.4 - 2.746 = 0.654$

21 X = number of robberies in a day $X \sim \text{Poisson}(4.32)$

a) $P(X=2) = \dfrac{e^{-4.32}(4.32)^2}{2!} = 0.1241$

b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X=0) + P(X=1)] = 1 - \left[\dfrac{e^{-4.32}(4.32)^0}{0!} + \dfrac{e^{-4.32}(4.32)^1}{1!}\right]$

$= 1 - 0.0708 = 0.9292$

22. X = weight of a baseball $X \sim \text{Unif}[5.085, 5.155]$

a) $P(X > 5.100) = (5.155 - 5.100)\left(1/(5.155 - 5.085)\right) = 0.7857$

b) $P(5.090 < X < 5.100) = (5.100 - 5.090)\left(1/(5.155 - 5.085)\right) = 0.1429$

c) $0.1 = (x - 5.085)\left(\dfrac{1}{5.155 - 5.085}\right) \longrightarrow x = 5.092$

23. $H_0: \mu = 14$ $H_a: \mu < 14$

Conditions: $n = 53 > 30$ σ unknown (use t)

Statistic: $t_0 = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{12 - 14}{4/\sqrt{53}} = -3.6401$ $df = 53 - 1 = 52$

p-value: $= 0.000314$ $pt(-3.6401, 52)$

Conclusion: p-value $< 0.01 = \alpha$. There is very strong evidence to suggest that the true mean fleece weight is not 14 lbs.

Interval: $\bar{x} \pm t_{\alpha/2, n-1} \, s/\sqrt{n} = 12 \pm 2.6778(4/\sqrt{53})$

$= 12 \pm 1.47 = (10.53, 13.47)$

We are 99% confident that the true mean fleece weight is between 10.53 and 13.47 pounds. Since this interval is entirely below the claimed value, we have reason to doubt the farmer's claim

24. $p_1 - p_2$ = difference in proportion of successfully removed warts with freezing (1) and duct tape (2) removal

$H_0: p_1 - p_2 = 0 \qquad H_a: p_1 - p_2 < 0$

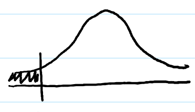
$\hat{p}_1 = 60/100 \qquad \hat{p}_2 = 83/104 \qquad \hat{p}_{pool} = \frac{60+83}{100+104} = \frac{143}{204}$

Conditions: $100\left(\frac{143}{204}\right) = 70.1 \qquad 100(61/204) = 29.9$

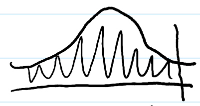
$104\,(143/204) = 72.9 \qquad 104(61/204) = 31.1$

Statistic: $$Z_0 = \frac{\left(\frac{60}{100} - \frac{83}{104}\right) - 0}{\sqrt{\frac{143}{204}\left(\frac{61}{204}\right)\left(\frac{1}{100} + \frac{1}{104}\right)}} = -3.09$$

p-value:



given



p-value < α So there is highly significant evidence to suggest the proportion of warts healed is lower with freezing than with duct tape.

99% confidence interval

$$\left(\frac{60}{100} - \frac{83}{104}\right) \pm 2.576\sqrt{\frac{\frac{60}{100} \times \frac{40}{100}}{100} + \frac{\frac{83}{104} \times \frac{21}{104}}{104}}$$

$= -0.198 \pm 2.576\,(0.04686) = -0.198 \pm 0.121$

$= (-0.319,\ -0.077)$

We are 99% confident the proportion of warts removed with duct tape is 7.7% to 31.9% higher than with freezing.