1)

Year	2013	2014	2015	2016	2017	2018
Rainfall	3.4	5.3	7.5	8.8	8.5	7.2

a) $\frac{3.4 + 5.3 + 7.5 + 8.8 + 8.5 + 7.2}{6} = \frac{40.7}{6} \boxed{= 6.7833}$

b) $S^2 = \frac{1}{6-1}\left[(3.4-6.78)^2 + (5.3-6.78)^2 + (7.5-6.78)^2 + (8.8-6.78)^2 + (8.5-6.78)^2 + (7.2-6.78)^2\right]$

$= \frac{21.348}{5}$

$= 4.2697$

$S = \sqrt{4.2697} \boxed{= 2.0663}$

c) Ordered rainfall: 3.4, 5.3, 7.2, 7.5, 8.5, 8.8

Median $\frac{7.2 + 7.5}{2} \boxed{= 7.35}$

Range: $\boxed{8.8 - 3.4 = 5.4}$

2) 3 real + 13 fake = 16 total

lose 5 P(no real lost)

P(1st fake) P(2nd fake) P(3rd fake) P(4th fake) P(5th fake)

$\left(\frac{\cancel{13}}{16}\right)\left(\frac{\cancel{12}}{15}\right)\left(\frac{11}{14}\right)\left(\frac{10}{\cancel{13}}\right)\left(\frac{9}{\cancel{12}}\right) = \frac{11 \times 10 \times 9}{16 \times 15 \times 14} = 0.2946$

3 $P(D) = 0.27$ $P(C) = 0.31$ $P(D \text{ and } C) = 0.22$

a) Population: People / people at UCR / UCR students

b) Each of the 50 randomly selected UCR undergrads

c) Preferred animals ; qualitative (categorical)

d) $P(C \text{ or } D) = P(C) + P(D) - P(C \text{ and } D)$

$= 0.31 + 0.27 - 0.22$

$\boxed{= 0.35}$

e) $P(C \text{ and } D) = 0.22$

$P(C)\,P(D) = 0.27 \times 0.31 = 0.0837$

$P(C \text{ and } D) \neq P(C)\,P(D)$ $\boxed{\text{NOT independent}}$

f) $P(C \text{ and } D) = 0.22 \neq 0$ so we can be in C and D simultaneously. So they are $\boxed{\text{NOT disjoint}}$

4) $P(M1) = 0.30$ $P(M2) = 0.39$ $P(M3) = 0.31$

$P(D|M1) = 0.04$ $P(D|M2) = 0.06$ $P(D|M3) = 0.02$

a) $P(D) = P(D \text{ and } M1) + P(D \text{ and } M2) + P(D \text{ and } M3)$

$= P(M1)\,P(D|M1) + P(M2)\,P(D|M2) + P(M3)\,P(D|M3)$

$= (0.3)(0.04) + (0.39)(0.06) + (0.31)(0.02)$

$= 0.012 + 0.0234 + 0.0062$

$\boxed{= 0.0416}$

b) $P(M2 \mid D) = \dfrac{P(M2 \text{ and } D)}{P(D)} = \dfrac{P(M2)\, P(D \mid M2)}{P(D)}$

$= \dfrac{(0.39)(0.06)}{0.0416} = \dfrac{0.0234}{0.0416} \boxed{= 0.5625}$

5) $H_0: p = 0.5$ (50% female)

$H_A: p \neq 0.5$ (not 50% female)

$\alpha = 0.05$ $\quad MoE \leq 0.02$

$0.02 \geq (Z_{\alpha/2})\left(\sqrt{\dfrac{p_0(1-p_0)}{n}}\right) = 1.96\sqrt{\dfrac{0.5(1-0.5)}{n}}$

$0.02 \geq 1.96\sqrt{\dfrac{0.25}{n}}$

$0.0102 \geq \sqrt{\dfrac{0.25}{n}}$

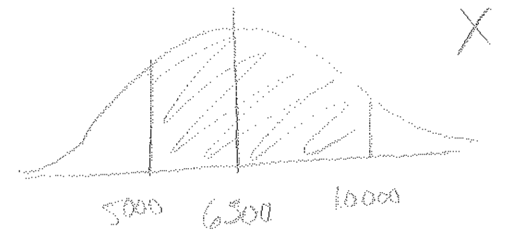
$1.041 \times 10^{-4} \geq \dfrac{0.25}{n}$

$n \geq \dfrac{0.25}{1.041 \times 10^{-4}}$

$\boxed{n \geq 2401 \text{ lizards.}}$

(4) X = avg monthly income in CA

$X \sim N(\mu = 6500,\ \sigma = 2100)$

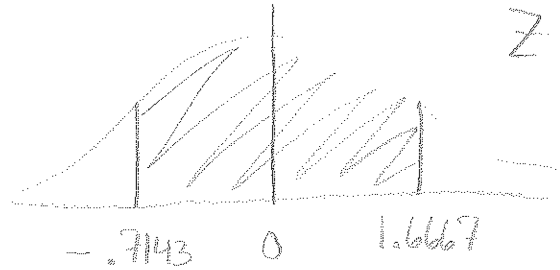


a.) $P(5{,}000 < X < 10{,}000)$

$$Z_{10{,}000} = \frac{10{,}000 - 6{,}500}{2{,}100} = 1.6667$$

$$Z_{5000} = \frac{5000 - 6500}{2100} = -0.7143$$

$$\Rightarrow P(-0.7143 < Z < 1.6667)$$

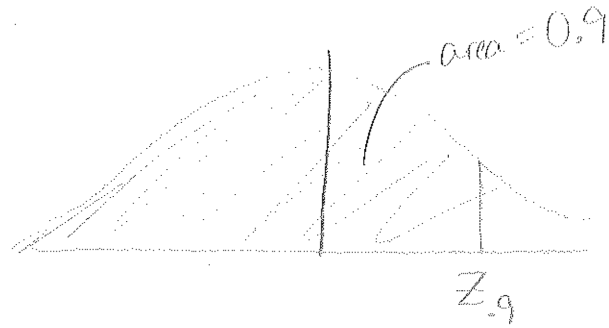


$$= 1 - P(Z > 1.6667) - P(Z < -0.7143)$$

$$= P(Z < 1.6667) - P(Z < -0.7143)$$

which is written fully in terms of left-tail probabilities for Z.

b) 90^{th} percentile?



$$0.9 = P(Z < z_{.9})$$

$$z_{.9} = \frac{6500 - x}{2100}$$

$$x = 6500 - 2100\, z_{.9}$$

7. X = times woken by cats

$X \sim \text{Pois}(\lambda = 2)$

a) $\mu = E(X) \boxed{= 2}$

$\sigma^2 = Var(X) = 2 \Rightarrow sd(X) = \sqrt{2} \boxed{= 1.4142}$

b) $P(X=0) = \dfrac{e^{-2} 2^0}{0!} = e^{-2} \boxed{= 0.1353}$

c) $P(X > 3) = 1 - P(X \leq 3)$

$= 1 - \left[P(X=0) + P(X=1) + P(X=2) + P(X=3)\right]$

$= 1 - \left[0.1353 + \dfrac{e^{-2} 2^1}{1!} + \dfrac{e^{-2} 2^2}{2!} + \dfrac{e^{-2} 2^3}{3!}\right]$

$= 1 - \left[0.1353 + 0.2707 + 0.2707 + 0.1804\right]$

$= 1 - 0.8571 \boxed{= 0.1429}$

8) $n = 1337$ 891 reported improvement $\rightarrow$ $\hat{p} = \frac{891}{1337} = 0.6664$

$H_0: p = 0.5$ (neither improve nor decline on average)

$H_A: p \neq 0.5$ (improve or decline on average)

a) Assume study data is taken well $\rightarrow$ independent ✓

Success-failure?

$np \approx n\hat{p} = 891 \geq 10$ ✓

$n(1-p) \approx n(1-\hat{p}) = 446 \geq 10$ ✓

$Z_{0.05/2} = 1.96$

$$SE = \sqrt{\frac{0.6664(1-0.6664)}{1337}} = 0.0129$$

$$\hat{p} \pm Z_{\alpha/2}(SE) \Rightarrow 0.6664 \pm 1.96(0.0129)$$

$$\Rightarrow 0.6664 \pm 0.0253$$

$$\Rightarrow (0.6411, 0.6917)$$

Since $p_0 = 0.5$ is NOT in the interval, there is sufficient evidence to reject H_0 and conclude that Vitamin C consumption tends to improve health outcomes at the $\alpha = 0.05$ level of significance.

b) $np_0 = n(1-p_0) = 668.5 \geq 10$ ✓

$\alpha = 0.01$

$\hat{p} = 0.6664$ $\quad\quad Z_{\alpha/2} = Z_{0.01/2} = 2.575$

$$ts = Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.6664 - 0.5}{\sqrt{\frac{0.5(0.5)}{1337}}}$$

$$= \frac{0.1664}{0.0137} = 12.1688$$

$$|ts| = |Z| = |12.1688| > |CV| = |Z_{0.01/2}| = |2.575|$$

Since the test statistic is more extreme than the critical value, there is sufficient evidence to reject H_0 and conclude that Vitamin C consumption tends to improve health outcomes at the $\alpha = 0.01$ level of significance

9) X = weight of male lab

$X \sim N(\mu = 71.5,\ \sigma = 7.5)$

$H_0: \mu = \mu_0 = 71.5$

$H_A: \mu \neq 71.5$

$n = 15 \quad \bar{X} = 80 \quad s = 5.1 \quad \alpha = 0.05$

$cv = Z_{\alpha/2} = 1.96$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{80 - 71.5}{7.5/\sqrt{15}} = \frac{8.5}{1.9365} = 4.3894$$

$$|ts| = |Z| = 4.3894 > 1.96 = |Z_{\alpha/2}| = |cv|$$

Since the test statistic is more extreme than the critical value, we reject H_0 and conclude that the true mean weight of male labrador retrievers is greater than 71.5 pounds.

10) a) Response: CD4-T cell count (y)
Predictor: # of cigarettes/day (x)

b)

x	8	6	5	3	10
y	202.4	427.1	219.3	529.19	111.7
$\hat{y}$	212.8753	317.9537	370.4929	475.5713	107.7969

$633.1889 - 52.5392(10)$

$$\hat{y} = 633.1889 - 52.5392x$$

residuals: $e_i = y_i - \hat{y}$

$y - \hat{y}$	-10.4753	109.1463	-151.1929	53.6187	3.9031

c) Predict for 2 cig/day $\rightarrow$ $x = 2$

$$\hat{y} = 633.1889 - 52.5392(2) = 528.1105$$

$x = 2$ is outside the range of our data so we are extrapolating a bit.

d) $$R = \frac{1}{n-1} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{s_x} \frac{y_i - \bar{y}}{s_y}$$

$\bar{x} = 6.4$ $\quad S_x = 2.7019$ $\quad \bar{y} = 297.9380$ $\quad S_y = 173.3179$

$x - \bar{x}$	1.6	-0.4	-1.4	-3.4	3.6
$y - \bar{y}$	-95.538	129.162	-78.638	231.252	-186.238

$\frac{x-\bar{x}}{s_x}$	0.5922	−0.1480	−0.5182	−1.2584	1.3324	
$\frac{y-\bar{y}}{s_y}$	−0.5512	0.7452	−0.4557	1.3343	−1.0745	
						Total
$\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$	−0.3264	−0.1103	0.2351	−1.6790	−1.4317	−3.3124

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right) = \frac{1}{4}(-3.3124) = \boxed{-0.8281}$$

There is a strong, negative linear relationship.

11)

X	0	1	2	3	Total
$P(X)$	0.614	0.196	0.128	0.062	
$X\,P(X)$	0	0.196	0.256	0.186	$\boxed{0.638 = E(X)}$
$X - E(X)$	−0.638	0.362	1.362	2.362	
$(X-E(X))^2$	0.4074	0.1310	1.8550	5.5790	
$p(x)[X-E(X)]^2$	0.2499	0.0257	0.2374	0.3459	$0.8590 = Var(X)$

$$sd(X) = \sqrt{Var(X)} = \sqrt{0.8590} = \boxed{0.9268}$$

b) Proportion over $E(X)$?

$E(X) = 0.638 \rightarrow$ Scores of 1, 2, 3 are over $E(x)$

$0.196 + 0.128 + 0.062 = 0.386$

OR $1 - 0.614 \boxed{= 0.386}$

c) $n = 10$ $Y = \#$ of people with depression

38.6 % of people have some level of depression (from b)

so $p = 0.386$

$\boxed{Y \sim Bin(n = 10, \; p = 0.386)}$

d) $E(Y) = \mu = np = 3.86$

$Var(Y) = \sigma^2 = np(1-p) = 2.3700$

e) $P(Y \text{ is more than } 8) = P(Y > 8) = P(Y=9) + P(Y=10)$

$= \binom{10}{9} 0.386^{9} \, 0.614^{10-9} + \binom{10}{10} 0.386^{10} \, 0.614^{10-10}$

$= 0.0012 + 0.000073$

$\boxed{= 0.0012}$