Chapter 1

1. Calculate the mean, the median and the variance / standard deviation of the following dataset on the observed dose of MDMA in 10 pills (in mg).

$$X = \{55,40,52,55,47,54,49,49,60,46\}$$

mean:

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

$$= \frac{55 + 40 + 52 + 55 + 47 + 54 + 49 + 49 + 60 + 46}{10}$$

$$= \frac{507}{10}$$

$$= 50.7$$

median: 40 46 47 49 49 52 54 55 55 60

$$\frac{49 + 52}{2} = \frac{101}{2} = 50.5$$

standard deviation:

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

$$= \frac{1}{9}((55 - 50.7)^2 + (40 - 50.7)^2 + (52 - 50.7)^2 + (55 - 50.7)^2 + (47 - 50.7)^2 + (54 - 50.7)^2 + (49 - 50.7)^2 + (49 - 50.7)^2 + (60 - 50.7)^2 + (46 - 50.7)^2)$$

$$= \sqrt{\frac{292.1}{9}} = 5.696978$$

1. Define the type of the following variables

a) The number of minutiae in friction ridge impressions:

Discrete (Quantitative)

b) The dose of MDMA in pills:

Continuous (Quantitative)

c) The design of the face of pills:

Nominal (Qualitative)

d) The size of shoes:

Ordinal (Qualitative)

e) Blood alcohol content:

Continuous (Quantitative)

f) The color of fibers:

Nominal (Qualitative)

g) The number of glass fragments transferred on a garment:

Discrete (Quantitative)

h) The size of garments:

Ordinal (Qualitative)

Chapter 2

1. The probability of observing an arch on any given person is 7%. The probability of observing a certain spatial arrangement of 4 minutiae is 8%. The probability to observe the same spatial arrangement of 4 minutiae on arches is **5%**.

a) Check if spatial arrangement and friction ridge pattern are independent

If independent …

$$P(A) * P(B) = P(A \cap B)$$

$$\begin{aligned} P(arch) * P(arrangement) &= 0.07 * 0.08 \\ &= 0.0056 \\ &\neq 0.05 \\ &= P(arch \ \cap arrangement) \end{aligned}$$

They are not ind□pendent.

b) Calculate the probability of observing the spatial arrangement given that you are looking at an arch.

$$\begin{aligned} P(arrangement|arch) &= \frac{P(arrangement \ \cap arch)}{P(arch)} \\ &= \frac{0.05}{0.07} \\ &= 0.7142857 \end{aligned}$$

c) Calculate the probability of observing something else than an arch

$$\begin{aligned} \square(arch') &= 1 - P(arch) \\ &= 1 - 0.07 \\ &= 0.93 \end{aligned}$$

d) Calculate the probability to observe the spatial arrangement on something else than an arch

There are two ways to solve this problem:

$$P(arrangement\ \cap arch') = P(arrangement|arch')P(arch')$$
$$= 0.03225 * 0.93$$
$$= 0.03$$

We get the probability of the arrangement, given that we have not observed an arch by considering the probability that we have not observed an arch (0.93), and by considering the percentage of individuals who have the arrangement, but do not have the arch (0.03 -> 0.08 have the arrangement, and 0.05 of that 0.08 have the arch, therefore, the remaining 0.03 have the arrangement, but do not have the arch)

$$P(arrangement\ \cap arch') = \ P(arch'|arrangement)P(arrangement)$$

$$Recall: P(arch'|arrangement) = 1 - P(arch|arrangement)$$

$$= 1 - \frac{P(arch\ \cap arrangement)}{P(arrangment)}$$

and so

$$P(arrangement\ \cap arch') = \left(1 - \frac{P(arch\ \cap arrangement)}{P(arrangment)}\right)P(arrangement)$$

$$= P(arrangement) - P(arch \cap arrangement)$$

$$= 0.08 - 0.05$$

$$= 0.03$$

e) Calculate the probability of observing an arch or the spatial arrangement

$$P(arch \cup arrangement) = P(arch) + P(arrangement) - P(arch \cap arrangement)$$
$$= 0.07 + 0.08 - 0.056$$
$$= 0.100$$

2. The probability to observe red viscose fibers on a garment is 3%.

   Note: We are looking at a GEOMETRIC distribution

a) Calculate the probability that we observe red viscose on the first garment we process

$$P(X = 1) = 0.03(1 - 0.03)^{1-1}$$
$$= 0.03$$

b) Calculate the probability that we observe red viscose on one of the first three garments that we process

$$P(X \leq 3) = P(X = 3) + P(X = 3) + P(X = 1)$$
$$= (1 - 0.03)^2 * (0.03) + (1 - 0.03)^1 * (0.03) + (1 - 0.03)^0 * (0.03)$$
$$= 0.087427$$

c) Calculate the probability that we need to process more than three garments to observe red viscose

$$P(X > 3) = 1 - P(X \leq 3)$$
$$= 1 - P(X = 3) - P(X = 2) - P(X = 1)$$
$$= 1 - 0.028227 - 0.0291 - 0.03$$
$$= 0.9126$$

3. The probability to observe a counterfeit penny is about 5%. We observe a sample of 100 pennies from a much larger population of pennies.

   Note: We are looking at a BINOMIAL distribution

a) Calculate the probability that we observe 4 (repeat for 5 and 6) counterfeit pennies in the sample of 100.

$$\begin{aligned}\Pr(X=4) &= \binom{100}{4}(1-0.05)^{96}0.05^{4}\\ &= 3921225 * 0.007268857 * 6.25e-06\\ &= 0.1781426\\ \Pr(X=5) &= \binom{100}{5}(1-0.05)^{95}0.05^{5}\\ &= 0.1800178\\ \Pr(X=6) &= \binom{100}{6}(1-0.05)^{94}0.05^{6}\\ &= 0.1500149\end{aligned}$$

b) Is the result in a surprising?

$$\textit{Not really, considering plots ob}\square\textit{erved in class, as well as that the largest percentage is occurring at } 5 \textit{ pennies, out of } 100\textit{, and our expected probability is } 5\%.$$

c) Calculate the probability to observe between 4 and 6 counterfeit pennies in the sample.

$$\begin{aligned}P(4 \le X \le 6) &= P(X=4) + P(X=5) + P(X=6)\\ &= 0.1781426 + 0.1800178 + 0.1500149\\ &= 0.5081753\end{aligned}$$

4. A sample of 100 white pills contains 60 pills composed of MDMA. You sample 50 pills out of the 100. What is the probability that 30 of them contain MDMA?

$$\Pr(X=30) = \frac{\binom{60}{30}\binom{100-60}{50-30}}{\binom{100}{50}} = 0.1615834$$

5. Solve the following equations:

Note: Use provided tables to solve for the following:

a) $P(Z \leq 2.58) = 0.99506$
b) $P(Z \leq -1.25) = 0.1056498$
c) $P(Z \geq 1.96) = 0.0249979$
d) $P(Z \leq z) = 0.7190$
$\Longrightarrow z = 0.5798734$

e) $P(T \geq t)_{df=12} = 0.01$
$\Longrightarrow t = 2.680998$
f) $P(T \leq t)_{df=18} = 0.995$
$\Longrightarrow t = 2.87844$

g) $P(X^2 \geq \chi^2)_{df=11} = 0.975$
$\Longrightarrow \chi^2 = 3.815748$
h) $P(X^2 \leq \chi^2)_{df=11} = 0.025$
$\Longrightarrow \chi^2 = 3.815748$
i) $P(X^2 \geq \square^2)_{df=10} = 0.01$
$\Longrightarrow \chi^2 = 23.20925$

6. Solve the following equation for $\mu = 15$ and $\sigma^2 = 4$.

$$\begin{aligned} P(X \leq 17) \Rightarrow Z &= \frac{17 - 15}{2} \\ &= \frac{2}{2} \\ &= 1 \\ &\Rightarrow P(Z \leq 1) = 0.5 + 0.3413 \\ &= 0.8414 \end{aligned}$$

$$\begin{aligned} P(X \geq 11.7) &\Rightarrow P(Z \geq \frac{11.7 - 15}{2} \\ &\Rightarrow P(Z \geq -1.65) \\ &= 0.950 \end{aligned}$$

$$\begin{aligned} P(12.5 \leq X \leq 16.5) &\Rightarrow P(-1.25 \leq Z \leq 0.75) \\ &= 0.3944 + 0.2734 \\ &= 0.667 \end{aligned}$$

Chapter 4

1. The purity of a shipment of 100 bags of cocaine is believed to be normally distributed. The purity of 10 bags has been measured.

x = {0.7599, 0.7582, 0.7291, 0.7475, 0.7530, 0.7482, 0.7596, 0.7705, 0.7434, 0.7410}

a) Estimate the purity of the shipment using a 95% confidence interval

$$\bar{X} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$$

$we\ have\ \ldots$

$$\bar{x} = 0.7483 \qquad s^2 = 0.0001388284 \qquad n = 10$$
$$t_{a/2} = -2.262157$$

$and\ so\ \ldots$

$$(0.7426593, 0.7595168)$$

b) What is the probability that the CI includes the true value of 0.75?

$$1.00$$

c) What would have happened if we were to analyze another 10 samples?

$The\ confidence\ interval\ itself\ would\ change:$
$We\ would\ have\ a\ new\ mean\ and\ standard\ deviation$
$correspondingto\ our\ new\ sample.$

2. A random sample of 100 individuals are tested for blood alcohol content. After having tested the first 30 individuals, it turns out that 12 of them have a BAC larger than the legal limit.

   a. Estimate the proportion of individuals that have a BAC larger than the legal limit using a 90% confidence interval.

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}}$$

$we\ have\ \ldots$

$$\hat{p} = 0.40 \qquad N = 100 \qquad n = 30$$

$such\ that\ \ldots$

$$0.40 \pm 1.65 \sqrt{\frac{0.40(0.60)}{30} * \frac{70}{99}}$$

$$(0.2759033, 0.5240967)$$

   b. Estimate the proportion of individuals that have a BAC larger than the legal limit using a 95% confidence interval.

$$0.40 \pm 1.96 \sqrt{\frac{0.40(0.60)}{30} * \frac{70}{99}}$$

$$(0.2525881, 0.5474119)$$

c. Repeat (a) and (b), knowing that 27 out of 60 individuals have a BAC larger than the legal limit.

$90\%\ Confidence\ Interval$:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}}$$

$we\ have\ \ldots$

$$\hat{p} = 0.45 \qquad N = 100 \qquad n = 60$$

$such\ that\ \ldots$

$$0.45 \pm 1.65 \sqrt{\frac{0.45(0.55)}{60} * \frac{40}{99}}$$

$$(0.382639, 0.517361)$$

$95\%\ Confidence\ Interval$:

$$0.45 \pm 1.96 \sqrt{\frac{0.45(0.55)}{60} * \frac{40}{99}}$$

$$(0.3699833, 0.5300167)$$

d. What can you observe by comparing (a), (b), and (c)?

$We\ see\ that, as\ expected, the\ 95\%\ CIs\ are\ wider\ than$
$the\ 90\%\ CIs. We\ also\ see\ the\ CIs\ tighten\ up\ when\ we\ consider$
$a\ greater\ proportion\ of\ the\ sample.$

Chapter 5

1. We want to characterize the proportion of individuals with arch friction ridge pattern in the general population.

    a. Calculate the sample size that we need to estimate that proportion with a precision of ± 0.01 and a confidence of 95%.

$$n = \frac{z_{\frac{\alpha}{2}}^2 \hat{p}(1-\hat{p})}{\epsilon^2}$$
$$= \frac{1.96^2 * 0.5 * 0.5}{0.01^2}$$
$$= 9604$$

    b. What would happen if you want to determine the same proportion (with the same precision and confidence) in a finite population of 1,000 people?

$$n = \frac{9604}{1 + \frac{9604 - 1}{1000}}$$
$$= 906$$

    c. What would happen if you want to redo (b), but you use the information that the proportion should be around 5%?

$$n = \frac{z_{\frac{\alpha}{2}}^2 \hat{p}(1-\hat{p})}{\epsilon^2}$$
$$= \frac{1.96^2 * 0.05 * 0.95}{0.01^2}$$
$$= 1825$$

$$n = \frac{1825}{1 + \frac{1825 - 1}{1000}}$$
$$= 647$$

We see a decrease i□ the necessary sample size.

Chapter 7

1. A study shows that 80 out of 120 fingerprint "identifications" were made based on more than 12 minutiae in common between the trace and control impressions. Test the hypothesis that more than 65% of "identifications" are made based on more than 12 minutiae.

   Hypotheses:

   $$H_0: p_0 = 0.65$$
   $$H_1: p_0 > 0.65$$

   We use a z-test for proportion:

   $$Z_{\alpha/2} = \frac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1 - p_0)}}$$
   $$= \frac{\sqrt{120}(0.6667 - 0.65)}{\sqrt{0.65 * 0.35s}}$$
   $$= 0.3827872$$

We can see right away that this value is very small, and thus close to zero. This will, intuitively, lead us to FAIL TO REJECT our null hypoth□sis.

Let's look at the $p -$ value, just to be certain.

$$p - value = 0.3509$$

Sure enough, our $p -$ value is large.

Fail to reject H_0: We do not have evidence to claim that th□ true proporti□n is not 65%.

2. Two garments are processed for foreign fibers. On the first garment, 190 foreign fibers (out of 336) are pink nylon, while on the second garment, 482 (out of 773) are pink nylon. Test whether the proportion of foreign pink nylon fibers is the same on both garments.

   Hypotheses:

$$H_0: p_1 = p_2 \;\; OR \;\; p_1 - p_2 = 0$$
$$H_1: p_1 \neq p_2 \; OR \; p_1 - p_2 \neq 0$$

$$For\ \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$z_{\alpha\backslash 2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$= \frac{(0.5654762 - 0.6235446)}{\sqrt{0.6059513 * 0.3940487 * \left(\frac{1}{336} + \frac{1}{773}\right)}}$$
$$= -1.818612$$

$$We\ can\ see\ that\ this\ vaue\ does\ not\ fall\ within\ our\ rejection\ region, whose\ threshold\ is\ 1.96$$

$$p - value = 0.0692$$

$$This\ is\ verified\ by\ our\ large\ p - value,$$

$$Fail\ to\ Reject\ H_0: We\ have\ sufficient\ evidence\ to\ claim\ that\ theproportions\ differ\ between\ garments.$$

”

3. The refractive indices of fragments from 2 different windows are compared to determine if the average refractive index of both windows is the same. Use the following data to perform the test:

$$n_a = 19;\ \bar{x}_a = 1.748421;\ s_a^2 = 0.579314$$
$$n_b = 28;\ \bar{x}_b = 1.386429;\ s_b^2 = 0.1651646$$

$$For\ S_{X_1X_2}^2 = \frac{(n_1-1)s_{x_1}^2 + (n_2-1)s_{x_2}^2}{n_1+n_2-2} = 0.3308 \qquad (by\ susbstitution)$$

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S_{X_1X_2}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$
$$= \frac{(1.748421 - 1.386429)}{\sqrt{0.3308\left(\frac{1}{19}+\frac{1}{28}\right)}}$$
$$= 2.117423$$

$$p - value = 1.96$$

$$Reject\ H_0: We\ have\ evidence\ to\ believe\ that\ the\ two\ windows\ have\ different\ average\ refractive\ indices.$$

4. A researcher is interested in comparing the rates of different shoe designs in different sub-populations. Test whether the distributions of patterns are different from one sub-population to another.

$$Observed\ Rates\ (O)$$

	Sport Shoes	City Shoes	Hiking Shoes	Casual Shoes	
Design A	56	83	43	55	237
Design B	25	44	18	11	98
Design C	23	53	21	33	130
Design D	45	89	38	60	232
Design E	28	37	17	17	99
	177	306	137	176	796

$$Expected\ Rates\ (E)$$

$$E = \frac{R_i C_j}{Total}$$

$$such\ that\ R_i\ and\ C_j\ are\ the\ row\ and\ column\ sumns, respectively, and\ Total\ is\ the\ total\ number\ of\ observation$$

	Sport Shoes	City Shoes	Hiking Shoes	Casual Shoes
Design A	52.7	91.1	40.6	52.4
Design B	21.8	37.7	16.9	21.7
Design C	28.9	50.0	22.4	28.7
Design D	51.6	89.2	39.9	51.3
Design E	22.0	38.1	17.0	21.9

$$Chi\ Squared\ Calculations$$

$$\frac{(E_k - O_k)^2}{E_k}$$

	Sport Shoes	City Shoes	Hiking Shoes	Casual Shoes
Design A	0.21	0.72	0.14	0.13
Design B	0.50	1.05	0.07	5.28
Design C	1.20	0.16	0.06	0.64
Design D	0.84	0.00	0.09	1.48
Design E	1.64	0.03	0.09	1.10

$$Chi\ Squared\ Test\ Statistic$$

$$\chi^2 = \sum_{k=1}^{n} \frac{(E_k - O_k)^2}{E_k}$$

$$= 15.43$$

$$answer\ may\ vary\ depending\ on\ rounding$$

$$Degrees\ of\ Freedom: (5-1)(4-1) = 4 * 3 = 12$$

$$Critical\ Value: 23.337$$

$$Fail\ to\ Reject\ H_0: We\ do\ not\ have\ evidence\ to\ believe\ that\ the\ distributions\ differ\ between\ sub-populations.$$

Chapter 11

1. A partial DNA profile is found at a crime scene and compared with that of Mr. X. The probability of observing the partial DNA profile at the crime scene given that the biological material was left by Mr. X is 0.67. The probability to observe the partial DNA profile if Mr. X is not the source of the biological material is 0.0001.

   a. Calculate the LR.

$$LR = \frac{P(observed\ DNA|Mr.X)}{P(observed\ DNA|Mr.X')}$$
$$= \frac{0.67}{0.0001}$$
$$= 6{,}700$$

   b. Calculate the probability that Mr. X is the source of the partial DNA profile if the population of the potential offender is 10,000.

$$\frac{1}{10{,}000} \times 6{,}700 = \frac{p}{1-p}$$
$$\Rightarrow .67 - .67p = p$$
$$\Rightarrow .67 = 1.67p$$
$$\Rightarrow p = \frac{.67}{1.67}$$
$$\Rightarrow p = 0.4012$$

   c. What would happen if it is 1,000,000?

$$\frac{1}{1{,}000{,}000} \times 6{,}700 = \frac{p}{1-p}$$
$$\Rightarrow .0067 - .0067p = p$$
$$\Rightarrow .0067 = 1.0067p$$
$$\Rightarrow p = \frac{.0067}{1.0067}$$
$$\Rightarrow p = 0.0067$$

   d. Does the LR change between (b) and (c)?

The LR does not change, however, the probabilities do change.

2. A finger impression is found at a crime scene and compared with a control impression from Mr. X by an examiner in laboratory A. The examiner declares that they “match”. Examiners of laboratory A are known to be very good at correctly declaring matches when the donors of the control impression are also the donors of the trace. Examiners from laboratory A are known to have an error rate of 1 in 100,000 cases.

    a. Calculate the LR.

$$
\begin{aligned}
LR &= \frac{P(observed\ print|Mr.X)}{P(observed\ print|Mr.\square')} \\
&= \frac{1}{\frac{1}{100{,}000}} \\
&= 100{,}000
\end{aligned}
$$

    b. Calculate the probability that Mr. X is the source of the trace if the population of potential offenders is 100,000.

$$
\begin{aligned}
\frac{1}{100{,}000} * 100{,}000 &= \frac{p}{1-p} \\
&\Rightarrow 1 - p = p \\
&\Rightarrow p = \frac{1}{2}
\end{aligned}
$$

    c. What would happen if one considers that police detectives propose the correct source (using non-fingerprint evidence) in about 80% of the cases?

$$
\begin{aligned}
100{,}000 \times \frac{0.8}{0.2} &= \frac{p}{1-p} \\
&\Rightarrow 400{,}000 - 400{,}000p = p \\
&\Rightarrow \frac{400{,}000}{400{,}001} = p \\
&\Rightarrow p = 0.99999
\end{aligned}
$$

d. What would happen if we assume prior odds that Mr. X is the source are 50/50?

$$100{,}000\times\frac{0.5}{0.5}=\frac{p}{1-p}$$

$$\Longrightarrow 100{,}000=\frac{p}{1-p}$$

$$\Longrightarrow 100{,}000-100{,}000p=p$$

$$\Longrightarrow \frac{100{,}000}{100{,}001}=p$$

$$\Longrightarrow p=0.99999$$

The probability remains the same.