Note # 4: Review of Chapters 1, 2, and 3

Problem 1. A study of fathers' involvement in their children's education interviews a random sample of fathers of school-aged children. One question concerns attendance at scheduled parent-teacher conferences. The table below shows the results:

	All	**Some**	**None**
Two-parent families	109	132	203
Single-parent families	15	10	13
Non-resident fathers	11	25	82

a. Create a contingency table that shows the distribution of attendance for each level of family structure.

b. Does it appear as though attendance and family structure are dependent or independent? Why?

Solution:

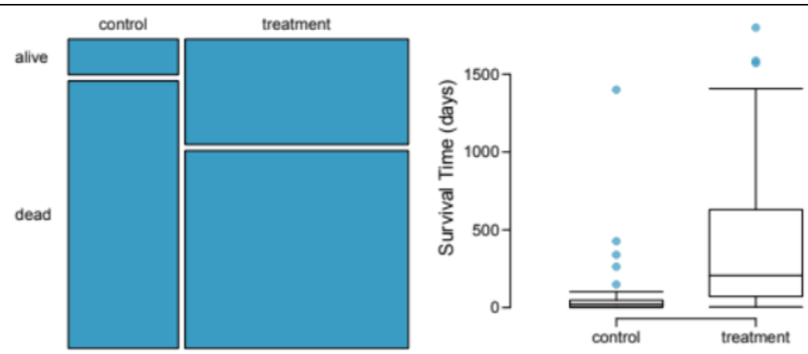
a.

Family structure	**All**	**Some**	**None**
Two-parent families	$\frac{109}{444} = 0.245$	$\frac{132}{444} = 0.297$	$\frac{203}{444} = 0.457$
Single-parent families	$\frac{15}{38} = 0.395$	$\frac{10}{38} = 0.263$	$\frac{13}{38} = 0.342$
Non-resident fathers	$\frac{11}{118} = 0.093$	$\frac{25}{118} = 0.212$	$\frac{82}{18} = 0.695$

b. If they were independent $\rightarrow P(all)$ the same for each group.

Since the proportions of those who attend all vs. some vs. more varies by family structure, they are dependent.

Problem 2. The Stanford University Heart Transplant Study was conducted to determine whether an experimental heart transplant program increased life span. Each patient entering the program was designated an official heart transplant candidate, meaning they were gravely ill and would most likely benefit from a new heart. Some patients got a transplant and some did not. The variable transplant indicates which group the patients were in; patients in the treatment group got a transplant and those in the control group did not. Another variable called survived was used to indicate whether or not the patient was alive at the end of the study. Of the 34 patients in the control group, 30 died. Of the 69 patients in the treatment group, 45 died. Researchers also measured the third variable for each patient, survival time, which recorded the
length of time each patient survived for.

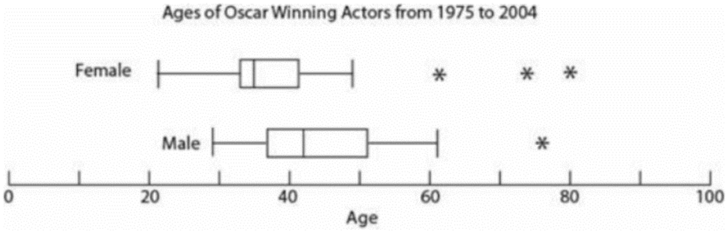


- **a.** What proportion of patients in the treatment group died?
- **b.** What proportion of patients in the control group died?
- **c.** Based on the mosaic plot, is survival independent of whether or not the patient got a transplant? Explain your reasoning.
- **d.** What do the box plots suggest about the efficacy (effectiveness) of the heart transplant treatment?

Solution:

- **a.** $P(died \mid trt.) = \frac{45}{69} = 0.652$
- **b.** $P(died \mid control) = \frac{30}{34} = 0.882$
- **c.** No, survival rates are different for treatment vs. control, they are dependent.
- **d.** While a large proportion of those in the treatment group died before the end of the study, on average their survival time was longer than those in the control group.

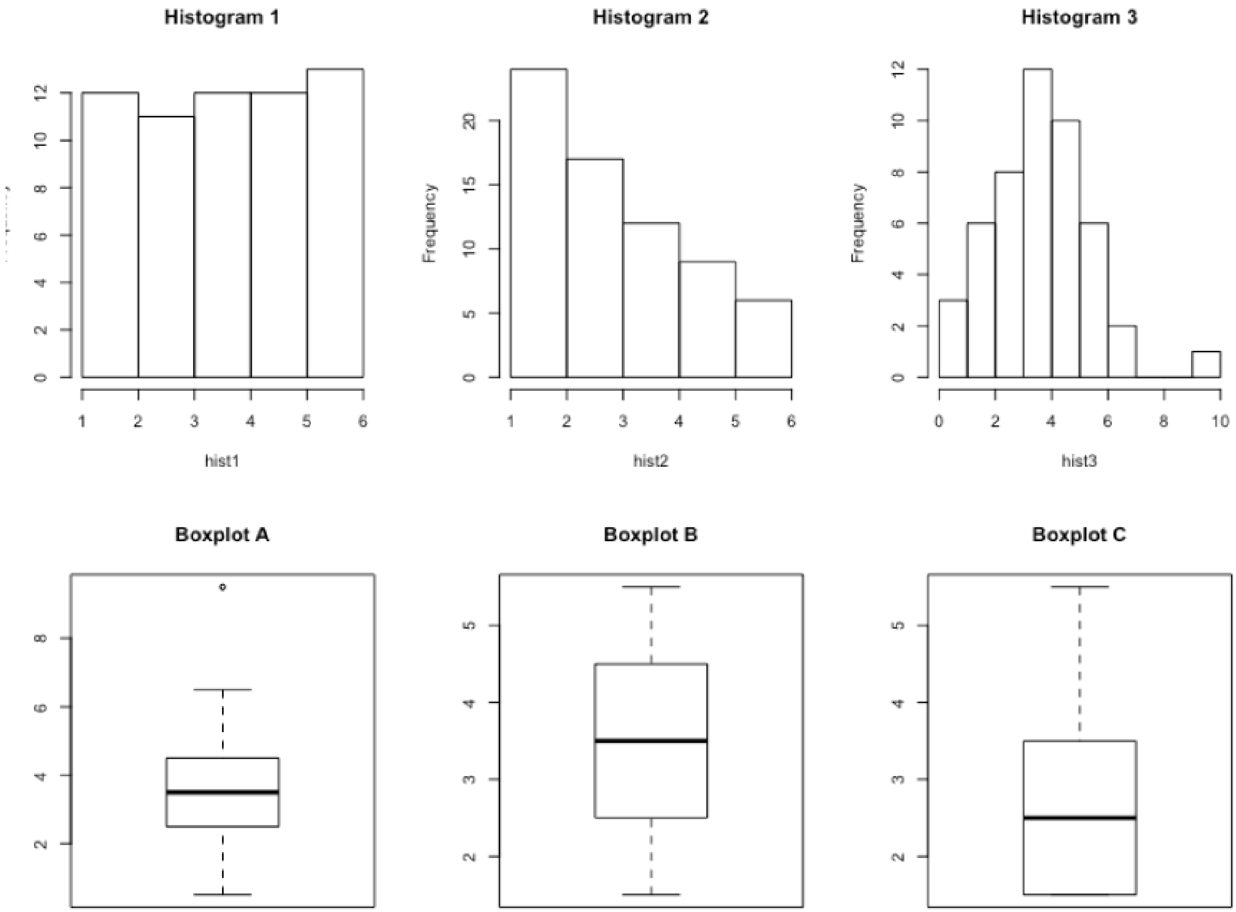
Problem 3. The following set of boxplots shows the 5-number summary for the ages of all Oscar Winning Actors from 1975 to 2004, split by gender. Based on these boxplots, does it appear that there is evidence that the typical age for males is different from the typical age for females?



Solution:

Maybe - the median ages for these two samples are different, however, their interquartile ranges (IQR) overlap quite a bit (possibly very similar).

Problem 4. Describe the distribution in the histograms below and match them to the box plots.



Solution:

Histogram 1: uniform, symmetric, no outliers → **Boxplot B**.

Histogram 2: unimodal, skewed right, no outliers → **Boxplot C**.

Histogram 3: unimodal, symmetric, potential outlier → **Boxplot A**.

Problem 5. The ***smallpox*** data set provides a sample of 6,224 individuals from the year 1721 who were exposed to smallpox in Boston. Doctors at the time believed that inoculation, which involves exposing a person to the disease in a controlled form, could reduce the likelihood of death. Each case represents one person with two variables: ***inoculated*** and ***result***. The variable ***inoculated*** takes two levels: ***yes*** or ***no***, indicating whether the person was inoculated or not. The variable ***result*** has two outcomes: ***lived*** or ***died***, indicating whether the person survived or not. This data set is summarized below.

	Yes	**No**	**Total**
Lived	238	5136	5374
Died	6	844	850
Total	244	5980	6224

a. What is the sample space?

b. What is the probability that a randomly selected individual survived?

c. What is the probability that a randomly selected individual who was inoculated survived?

d. What is the probability that a randomly selected individual who was not inoculated survived?

e. Does it seem like inoculation and survival are independent?

Solution:

a. S= {(inoculated, lived), (inoculated, died), (not inoculated, lived), (not inoculated, died)}

b. $P(survived) = \frac{\#\, survived}{total\, \#} = \frac{5374}{6224} = 0.8634$

c. $P(survived|\text{inoculated}) = \frac{\#\, survived\ and\ \text{inoculated}}{\#\,\text{inoculated}} = \frac{238}{244} = 0.9754$

d. $P(survived|not\ \text{inoculated}) = \frac{\#\, survived\ and\ not\ \text{inoculated}}{\#\, not\ \text{inoculated}} = \frac{5136}{2445980} = 0.8589$

e. $P(survived)$ does differ based on inoculation status → we believe they are dependent

Problem 6. A 2010 **SurveyUSA** poll asked 500 Los Angeles residents, "What is the best hamburger place in Southern California? Five Guys Burgers? In-N-Out Burger? Fat Burger? Tommy's Hamburgers? Umami Burger? Or somewhere else?" The distribution of responses by gender is shown below.

	Male	Female	Total
Five Guys Burgers	5	6	11
In-N-Out Burger	162	181	343
Fat Burger	10	12	22
Tommy's Hamburgers	27	27	54
Umami Burger	5	1	6
Other	26	20	46
Not Sure	13	5	18
Total	248	252	500

a. Are being female and liking In-N-Out Burger best mutually exclusive?

b. What is the probability that a randomly selected male likes In-N-Out the best?

c. What is the probability that a randomly selected female likes In-N-Out the best?

d. What is the probability that a man and a woman who are dating both like In-N-Out the best? Note any assumptions you make and evaluated whether you think they are reasonable.

e. What is the probability that a randomly selected person like In-N-Out best or that person is female?

Solution:

a. No, 181females picked In-N-Out.

b. $P(\text{INO}|\text{male}\ ) = \frac{\#\, males\ and\ \text{INO}}{\#\ \text{males}} = \frac{162}{248} = 0.65$

c. $P(\text{INO}|\text{female}\ ) = \frac{\#\, females\ and\ \text{INO}}{\#\ \text{females}} = \frac{181}{252} = 0.72$

d. $P(\text{male INO} \cap female\ INO) = \ P(\text{male INO}) \times P(\ female\ INO) = 0.65 \times 0.72 = 0.468$

assume two people were independent valid? Maybe, don’t know for sure.

e. $P(\text{INO} \cup female) = \frac{\#INO + \#\, females - (\#INO\ and\ female}{total\ 3} = \frac{343+252-181}{500} = \frac{414}{500} = 0.828$

Problem 7. Many times, when we are discussing probabilities about diseases, we talk about the risk and the odds. The risk is the same as the probability, while the odds is the probability divided by one minus the probability. One of the most common genetic disorders in the United States is Down Syndrome. If an individual has Down Syndrome, there is an increased chance that they will have a heart defect. Approximately 47% of infants born with Down Syndrome also have a heart defect?

- **a.** If a child has Down Syndrome, what is the risk of them having a heart defect?
- **b.** If a child has Down Syndrome, what is the odds of them having a heart defect? What does this mean?

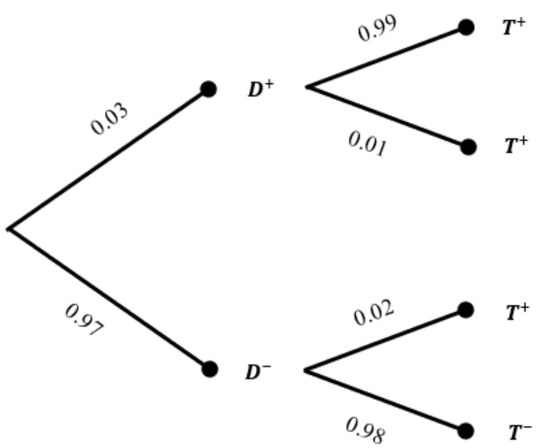
Solution:

- **a.** $P(\text{Heart Defect}|\text{Down Syndrome}) = 0.47$
- **b.** $Odds = \frac{P}{1-P} = \frac{0.47}{1-0.47} = \frac{0.47}{0.53} = 0.887$. Children with Down Syndrome are more likely to not have a heart defect.

Problem 8. A genetic test is used to determine if people have a predisposition for thrombosis, which is the formation of a blood clot inside a blood vessel that obstructs the flow of blood through the circulatory system. It is believed that 3% of people actually have this predisposition. The genetic test is 99% accurate if a person actually has the predisposition, meaning that the probability of a positive test result when a person actually has the predisposition is 0.99. The test is 98% accurate if a person does not have the predisposition. What is the probability that a randomly selected person who tests positive for the predisposition actually has the predisposition?

Solution:

$$prev = \boldsymbol{P(D^+)} = 0.03 \text{,} \quad P(T^+ | D^+) = 0.99 \text{,} \quad P(T^- | D^-) = 0.98$$



$$P(D^+ \cap T^+) = 0.03 \times 0.99 = 0.0297$$

$$P(D^- \cap T^+) = 0.97 \times 0.02 = 0.0194$$

$$P(D^+ | T^+) = \frac{P(D^+ \cap T^+)}{P(T^+)} = \frac{0.0297}{0.0297 + 0.0194} = 0.6049$$

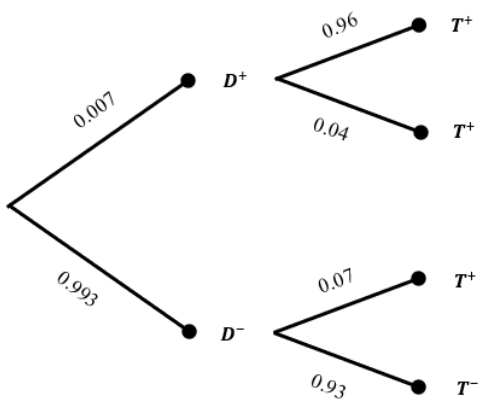
Problem 9. Based on past drug testing of air traffic controllers, the FAA reports that the probability of drug use at any given time is approximately 0.007. The FAA uses a particular test to determine if the air traffic controllers are currently using drugs is 96% sensitive and 93% specific

a. What is the probability of a positive test?

b. If a test is positive, what is the probability that the individual is actually using drugs?

Solution:

$$prev = \boldsymbol{P(D^+)} = 0.007 \text{ , } \boldsymbol{Sens} = P(T^+|\, D^+) = 0.96 \text{ , } \quad \boldsymbol{Spec} = P(T^-|\, D^-) = 0.93$$



$P(D^+ \cap T^+) = 0.007 \times 0.96 = 0.00672$

$P(D^- \cap T^+) = 0.993 \times 0.07 = 0.06951$

a. $P(T^+) = P(T^+ \cap D^+) + P(T^+ \cap D^-) = 0.00672 + 0.06951 = 0.07623$

b. $P(D^+|\, T^+) = \frac{P(D^+ \cap T^+)}{P(T^+)} = \frac{0.00672}{0.07623} = 0.0882$