**2.34 Card game. Consider the following card game with a well-shuffled deck of cards. If you draw a red card, you win nothing. If you get a spade, you win \$5. For any club, you win \$10 plus an extra \$20 for the ace of clubs.**

**(a) Create a probability model for the amount you win at this game. Also, find the expected winnings for a single game and the standard deviation of the winnings.**

We know that we have 4 possible outcomes:

Outcome 1: Draw a red card (half the deck, or probability 0.50) — \$0 payout

Outcome 2: Draw a spade (one quarter of the deck, or probability 0.25) — \$5 payout

Outcome 3: Draw 2-K of clubs (12 out of 52 cards, or probability 0.23) — \$10 payout

Outcome 4: Draw A of clubs (1 out of 52 cards, or probability 0.02) — \$30 payout

Note that this accounts for all cards and a probability summing 1.0.

We can begin to represent the probability model by creating a table with the various possibilities:

| $i$ | Any Red Card | Any Spade | 2-K of Clubs | Ace of Clubs | Total |
|---|---|---|---|---|---|
| $x_i$ | \$0 | \$5 | \$10 | \$30 | |
| $P(X = x_i)$ | 0.50 | 0.25 | 0.23 | 0.02 | 1.00 |

We will calculate the expected value — E(X) — for this game, in which the random variable X represents the card I as a gambler pull and x represents the various possibilities. We can determine the expected value with the equation

$$E(X) = \sum_{i=1}^{4} x_i P(X = x_i)$$

I sum the products of each outcome's probability and payout (the product of each column's values) to find the expected value:

(0)(0.50) +(5)(0.25) + (10)(0.23) + (30)(0.02)

0 + 1.25 + 2.3 + 0.60

4.15

The expected value is \$4.15 — that's about what I can expect to win on average for any given hand. Note that this only makes sense if I'm playing a few hands — there's no way to win \$4.15 on just one or two plays… it's an average which assumes I keep gambling.

To find the standard deviation, I can continue filling out my probability model:

| $i$ | **Any Red Card** | **Any Spade** | **2-K of Clubs** | **Ace of Clubs** | **Total** |
|---|---|---|---|---|---|
| $x_i$ | \$0 | \$5 | \$10 | \$30 | |
| $P(X = x_i)$ | 0.50 | 0.25 | 0.23 | 0.02 | 1.00 |
| We multiply the two rows above in each column to find out what each possibility contributes to the expected value / population mean, here in **RED…** | | | | | |
| $x_i * P(X = x_i)$ | 0 | 1.25 | 2.30 | 0.60 | **4.15** |
| Then we find the difference between $x_i$ and the expected value in the next row: | | | | | |
| $x_i$ - $E(X)$ | -4.15 | 0.85 | 5.85 | 25.85 | |
| I square each difference and multiply it by its probability to get each possibility's contribution to the variance, here in **BLUE…** | | | | | |
| $(x_i - E(X))^2 * P(X = x_i)$ | 8.611 | 0.1806 | 7.87 | 13.364 | **30.03** |

The square root of my variance( = 30.03), aka the standard deviation, is 5.48.

I can also solve this empirically, by using the counts of the various cards:

The standard deviation can be calculated by taking the square root of the average of the squared differences between the expected value and the actual value.

We have 26 examples of winning nothing at all (red cards), so the squared difference for each of those cases would be $(4.15)^2$.

We have 13 examples of winning \$5, so the squared difference for each of those cases would be $(0.85)^2$.

We have 12 examples of winning \$10, so in these cases, the squared difference would be $(5.85)^2$.

We have one example of winning \$30, with a squared difference of $(25.85)^2$.

Adding these squared differences together, we get

$(26)(4.15)^2 + (13)(0.85)^2 + (12)(5.85)^2 + (1)(25.85)^2$

$447.785 + 9.93925 + 410.67 + 668.2225$

1536.617 is the sum of my squared differences.  I divide by 51 (n-1 for an unbiased standard deviation) to get my variance, the average of the squared distance.  In this case,  30.13  is my variance, and the square root of that, 5.48, is the standard deviation.

**b)  What is the maximum amount you would be willing to pay to play this game? Explain.**

I enjoy gambling, but I don't want to enrich the house more than necessary, so, assuming the game were to continue with various draws / attempts, I'd pick a break-even point as my ceiling for the amount I'd bet.  Betting $4.15 per hand means I'll tend to break even over the long haul.

**2.40 Baggage fees. An airline charges the following baggage fees: $25 for the first bag and $35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.**

**(a)  Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.**

We have three possible outcomes:

Outcome 1: Passenger has 0 bags.  Revenue = 0, probability = 0.54

Outcome 2: Passenger has 1 bag.  Revenue = $25, probability = 0.34

Outcome 3: Passenger has 2 bags.  Revenue = $60, probability = 0.12

Again we can build a table that constitutes a probability model:

| $i$ | 0 bags | 1 bag | 2 bags | Total |
|---|---|---|---|---|
| $x_i$ | $0 | $25 | $60 | |
| $P(X = x_i)$ | 0.54 | 0.34 | 0.12 | 1.00 |
| We multiply the two rows above in each column to find out what each possibility contributes to the expected value / population mean, here in **RED…** | | | | |
| $x_i * P(X = x_i)$ | 0 | 8.5 | 7.2 | **15.7** |
| Then we find the difference between $x_i$ and the expected value in the next row: | | | | |
| $x_i$ - $E(X)$ | -15.7 | 9.3 | 44.3 | |
| I square each difference and multiply it by its probability to get each possibility's contribution to the variance, here in **BLUE…** | | | | |
| $(x_i - E(X))^2 * P(X = x_i)$ | 133.10 | 29.41 | 235.50 | **398.01** |

The average revenue per passenger is $15.70 and the population standard deviation is the square root of 398.01, or $19.95.

**(b) About how much revenue should the airline expect for a flight of 120 passengers? With what standard deviation? Note any assumptions you make and if you think they are justified.**

The expected revenue would be 120 * the average revenue per passenger ($15.70), if we assume that the 120 passengers are a truly random sample of the airline passenger population. Therefore the airline could expect revenue of $1884.00. The standard deviation can be calculated by taking the variance for each passenger (398.01), multiplying it by the number of passengers, and taking the square root. The total variance is 398.01 * 120, or 47761.20. Taking the square root of that gives me 218.54. The standard deviation is $218.54.

Note that I'm assuming that this flight of 120 passengers is a good representative sample of the population, which is a hard assumption to swallow. Any flight contains passengers going to the same destination, which means that the pattern of baggage will be more likely to be similar. Let's say this flight is a small plane going to Cuba, with most passengers loading up 2 bags apiece, each stuffed with a multitude of items that are hard to obtain on the island. Or take the example of a large commuter flight between Buffalo and NYC, which likely contains business travelers, most of whom have only a purse or briefcase and no checked bags at all. The 120 flyers do not constitute a random sample of independent variables and so the assumptions underlying my calculations are most likely not justified.

2.42 Selling on Ebay. Marcie has been tracking the following two items on Ebay:

**• A textbook that sells for an average of $110 with a standard deviation of $4.**

**• Mario Kart for the Nintendo Wii, which sells for an average of $38 with a standard deviation of $5.**

**(a) Marcie wants to sell the video game and buy the textbook. How much net money (profits - losses) would she expect to make or spend? Also compute the standard deviation of how much she would make or spend.**

Let's note the particulars of each item (sans units)

Textbook — E(T) = 110; standard deviation = 4; variance = 16

Game — E(G) = 38; standard deviation = 5; variance = 25

We have a linear combination of variables here, so Marcie's trading income can be modeled by 1G + -1T (one game - one textbook), or 38 - 110. She can expect to end up with a net change of -72, or overall spending of $72.

The standard deviation of Marcie's commerce can be modeled by $\sqrt{\text{Var(G) -Var(T)}}$ , or $\sqrt{25-16}$ , which equals 3. The standard deviation is $3.

**(b) Lucy is selling the textbook on Ebay for a friend, and her friend is giving her a 10% commission (Lucy keeps 10% of the revenue). How much money should she expect to make? With what standard deviation?**

In Lucy's case, the coefficient of T is 0.10, so the amount of money she can expect to make is 0.10 * 110, or $11. For the standard deviation, we square the coefficient and multiply that by the variation of the textbook, then take the square root of that: $\sqrt{(0.10^2 * 16)}$, or 0.4. The standard deviation for Lucy's commission is 40 cents.

2.46 Income and gender. **The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 59% males and 41% females.**

| *Income* | *Total* |
|---|---|
| $1 to $9,999 or loss | 2.2% |
| $10,000 to $14,999 | 4.7% |
| $15,000 to $24,999 | 15.8% |
| $25,000 to $34,999 | 18.3% |
| $35,000 to $49,999 | 21.2% |
| $50,000 to $64,999 | 13.9% |
| $65,000 to $74,999 | 5.8% |
| $75,000 to $99,999 | 8.4% |
| $100,000 or more | 9.7% |

**(a) Describe the distribution of total personal income.**

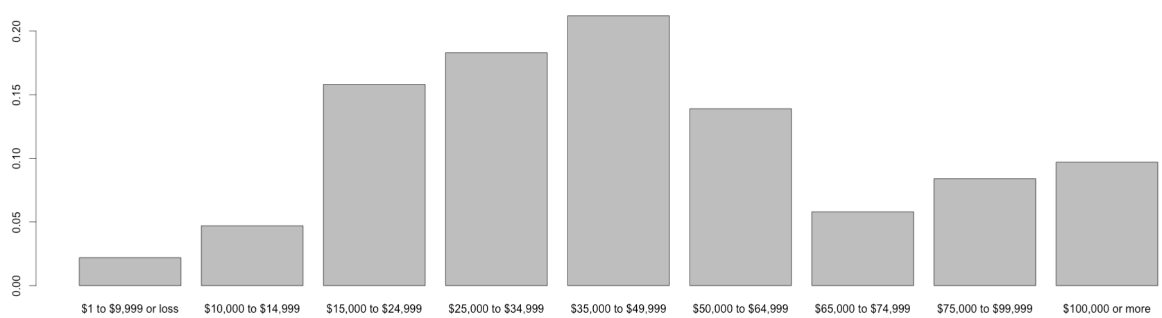
Since we already have probabilities for each bin, we will use boxplot() instead of hist() to visualize this dataset in R.

```
income<-(c("$1 to $9,999 or loss", "$10,000 to $14,999", "$15,000 to
$24,999","$25,000 to $34,999", "$35,000 to $49,999", "$50,000 to
$64,999", "$65,000 to $74,999","$75,000 to $99,999", "$100,000 or
more"))

rates<-c(0.022, 0.047, 0.158, 0.183, 0.212, 0.139, 0.058, 0.084,0.097)

income_data<-data.frame(income,rates)

barplot(income_data$rates, names=income_data$income)
```




The mode or most populated bin for income is $35,000-$49,999, and the distribution looks like a normal distribution with a bloated upper tail. Very few people are in the lower extremes, most people have income in the middle categories, and a few (but more than in the lower extremes) have incomes in the higher extremes. Because the size of the bin varies in the income categories (from a low of $5000 in the $10,000-$14,999 category to an undisclosed size in the first and last category, we don't have the granular information that would help us draw additional, firmer conclusions.

**(b) What is the probability that a randomly chosen US resident makes less than $50,000 per year?**

Since the data we were given comes from a "representative sample" and this represents an "or" conjunction of probabilities, we can simply add together the various probabilities. The probability that a randomly chosen US resident makes less that $50,000 per year is 2.2% + 4.7% + 15.8% + 18.3% + 21.2%, or 62.2%.

**(c) What is the probability that a randomly chosen US resident makes less than $50,000 per year and is female?**

**Note any assumptions you make.**

Again, we are told that the instrument uses a representative sample, something which surprises a bit, given the skew towards males, but let's assume that the sample proportion of 59% male and 41% female is a good proxy for the population distribution.

Our data gives us two probabilities: the probability of being female is 0.41, and the probability of having an income of less than $50000 is 0.622. We assume that these two factors are independent, since we have no granular information about income and sex together. Therefore we multiply these probabilities to get 0.255, or 25.5%.

**(d) The same data source indicates that 71.8% of females make less than $50,000 per year. Use this value to determine whether or not the assumption you made in part (c) is valid.**

When I compare my answer to (b), 62.2%, to the reality of 71.8% female income below 50K, clearly, my calculation is far away from the truth. This means that the female income distribution is not the same as the overall distribution (and by extension, the male distribution, which is the only other factor included in overall distribution). Women are disproportionately represented in the lower income categories, which indicates that men tend to make more than women. The assumption I made in part c is invalid.