Lecture 13
Last time
Last time

• Sampling with and without replacement

• Review of the binomial distribution

• Expectation
Expectation

• Given a random variable $X$ with probability distribution $p(x) = \text{pr}(X=x)$

• The expected value of $X$ is defined to be

$$E(X) = \sum_{\text{values } x} x \cdot \text{pr}(x)$$

• If $X$ represents your “winnings” in a game of chance, then $E(X)$ is your expected payout
Population mean

• We also refer to $E(X)$ as the population mean

$$
\mu = E(X) = \sum_{\text{values } x} x \times \text{pr}(x)
$$
Two rules for means

• \( E(aX + b) = aE(X) + b \)

• Let \( X_1 \) and \( X_2 \) be any two random variables

\[ E(X_1 + X_2) = E(X_1) + E(X_2) \]
Population variance

- The sample standard deviation represented the “spread” in the data; the population standard deviation does the same for the population.

- The population variance is given by

\[ \text{var}(X) = \sigma^2 = \text{E}(X-\mu)^2 = \sum_{\text{values } x} (x-\mu)^2 \text{ pr}(x) \]

- And the population standard deviation is

\[ \text{sd}(X) = \sqrt{\sigma^2} \]
Population variance

• Measures the spread of the distribution, or how widely X varies

Those who stop their inquiries at the mean have souls “as dull to the charm of variety as that of a native of one of our English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once.”

Francis Galton (1889)
Rules for variances

- If \( a \) and \( b \) are constants

\[
\text{var}(aX + b) = a^2 \text{var}(X) \quad \text{or} \quad \text{sd}(aX+b) = |a| \text{sd}(X)
\]
Rules for variances

• Let $X_1$ and $X_2$ be two random variables defined on independent trials

• Addition rule for variances

$$\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2)$$
Question

• Let $X_1$ and $X_2$ be two random variables defined on independent trials

• What is $\text{var}(X_1 - X_2)$?
An example

• Suppose we conduct an experiment and create a random variable $X$ that is 1 if a certain event occurs and 0 otherwise

• $\text{pr}(X=1) = \text{pr}( \text{the event occurs } ) = p$

• $\text{pr}(X=0) = \text{pr}( \text{the event doesn’t occur} ) = 1-p$
An example

- Recall that the mean is

$$E(X) = 1 \times \text{pr}(X=1) + 0 \times \text{pr}(X=0)$$

- So that the variance is

$$(E(X - p)^2 = (1-p)^2 \times \text{pr}(X = 1) + p^2 \times \text{pr}(X = 0)$$

$$= (1-p)^2p + p^2(1-p)$$

$$= p(1-p)$$
Expectation

• Suppose we conduct an experiment $n$ times

• Each time we check to see if an outcome occurs

• We create $n$ random variables $X_1, \ldots, X_n$ so that $X_k$ is $1$ if the outcome occurs on trial $k$

• $X_1 + X_2 + \ldots + X_n$ has a binomial distribution
Expectation of the binomial

• The expected number of trials having the outcome is

\[ E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) \]

\[ = p + p + \ldots + p \]

\[ = np \]
Variance of the binomial

• Since the trials are independent

\[ \text{var}(X_1 + X_2 + \ldots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \ldots + \text{var}(X_n) \]

\[ = p(1-p) + p(1-p) + \ldots + p(1-p) \]

\[ = n \ p \ (1-p) \]
Summary

• The expected value of a binomial distribution with sample size $n$ and success probability $p$ is $np$.

• The variance of a binomial distribution with sample size $n$ and success probability $p$ is $np(1-p)$. 
Continuous distributions

- Recall our friend the histogram
- The height of each bar is the number of data points that fall into that interval
Continuous distributions

- We *standardize* a histogram by rescaling the heights so that the *area* of each bar is equal to the relative frequency of data points falling into each interval.

- Assume that each bar has width $w$ and out of $n$ points, let $f_k$ be the frequency of points in the $k$th interval.

- Then, $f_k/n$ is the relative frequency, and the standardized histogram has height $f_k/(nw)$. 
Each bar is 1,000 wide, and its height is the relative frequency divided by the interval width.
Continuous distributions

• The area of the kth bar is now just the proportion of data points falling into the kth interval.

• We can now describe any event involving the intervals and compute its probability by taking the area under the (standardized histogram).

• In particular, the total area under the histogram is 1.
pr( X ≥ 30,000) = 0.64
Continuous distributions

• A continuous random variable $X$ is described by a density function or curve $p(x)$

• Probabilities of events are computed as areas under the density curve

• The total area under the curve is 1
Lecture 14
Last time

- Review of expectation
- Population mean and variance
- Example: The binomial distribution
- Continuous random variables
Recall the random variable

• A random variable $X$ is a measurement taken on the outcome of a random experiment

• We assign probabilities to events involving $X$ using a probability function $p(x)$

• $p(x)$ represents a mathematical model
$X = \text{the number of H's in 10 coin tosses}$
Discrete random variables

- X can only take discrete values

- To compute the probability of an event we add p(x) for all x that make up the event

- \(0 \leq p(x) \leq 1\) and \(\sum \text{values } x p(x) = 1\)
Discrete random variables

- We can think of $p(5)$ as the long-run relative frequency of the event “$X=4$” or “$X \geq 8$”

- That is, we repeat our trial many, many times and look at the proportion in which the event occurs
Relative frequency of the event “X=4” over 5,000 trials
Relative frequency of the event “\(X \geq 8\)” over 5,000 trials
Discrete random variables

- $p(x)$ is a mathematical model from which we draw samples

- Of course, there will be sampling variability and hence our relative frequencies appear noisy

- Keep clear the distinction between a model $p(x)$ that describes our population, and relative frequencies that we compute from a sample
Discrete random variables

• In your lab, you considered two events

   A = “An ant is 50 cm north of the nest at t=10,000”
   B = “An ant is 50 cm east of the nest at t=10,000”

• Under our probability model A and B are independent
  and \( \text{pr}(A \text{ and } B) = \text{pr}(A) \text{ pr}(B) \)

• Watch a sample of ants and compute the proportion
  that satisfy A, B

• Due to sampling variability, we will rarely see the
  product of these proportions equaling the proportion
  of our ants satisfying “A and B”
Continuous random variables

• X can now take continuous values

• We assign probabilities to events involving X using a probability density function $p(x)$

• $p(x)$ represents a mathematical model
p(x) for a continuous random variable X
Continuous random variables

- We compute the probability of events using the area under the density curve

- \( p(x) \geq 0 \) and \( p(x) \) is normalized so the area under the entire curve is 1
\text{pr}( 8 \leq X \leq 10)
\[ p( X \geq 8) \]
Continuous random variables

• We can again think of \( \text{pr}(X \geq 8) \) as the long-run relative frequency of the event “\( X \geq 8 \)”

• That is, we repeat our trial many, many times and look at the proportion in which the event occurs
Relative frequency of the event “$8 \leq X \leq 10$” over 5,000 trials
Relative frequency of the event “$X \geq 8$” over 5,000 trials
Continuous random variables

• $p(x)$ is a mathematical model from which we draw samples

• Of course, there will be sampling variability and hence our relative frequencies appear noisy

• Keep clear the distinction between a model $p(x)$ that describes our population, and relative frequencies that we compute from a sample
1,000 samples
10,000 samples
Continuous random variables

• Because our probabilities are based on the area under $p(x)$, the chance that $X$ takes on any single value is zero

• Because of this, we usually focus on events that can be expressed in terms of some collection of intervals
lower tail, $\Pr(X \leq x)$

upper tail, $\Pr(X \geq x)$
Cumulative distribution function

• For each $x$, we plot $\Pr(X \leq x)$

• Or, for each $x$, we plot the value of the lower tail

• It increases from 0 to 1 as $x$ increases
Cumulative distribution function
Quantiles

• The p-quantile is defined to be the point $x_p$ such that

$$\text{pr}(X \leq x_p) = p$$

• We can read this from the cumulative distribution function
The 0.25 and 0.75 quantiles (also known as the 25th and 75th percentiles)
The normal distribution

• The bell-shaped curve

• Commonly referred to as the “law of error”

\[ p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \]

• \( E(X) = \mu \), and \( \text{var}(X) = \sigma^2 \)
Several normal curves

\[
\begin{align*}
\mu &= 10, \quad \sigma^2 = 9 \\
\mu &= 25, \quad \sigma^2 = 49 \\
\mu &= 40, \quad \sigma^2 = 1
\end{align*}
\]
The *standard* normal distribution has mean 0 and variance 1.
Standard units

• If $X$ has a standard normal distribution, then $aX+b$ again has a normal distribution but with mean $b$ and variance $a^2$

• Similarly, if $X$ has mean $\mu$ and variance $\sigma^2$, then $\frac{X-\mu}{\sigma}$ has a standard normal distribution
Male heights from the BRFSS data
Female heights from the BRFSS data
One deck of cards
With (black) and without (cyan) replacement
Two decks of cards
With (black) and without (cyan) replacement
Five decks of cards
With (black) and without (cyan) replacement

![Graph showing the probability of drawing a certain number of red cards from five decks of cards with and without replacement.](image)
Lecture 15
Last time

• Continuous random variables

• Cumulative distribution function and quantiles

• The normal distribution
The normal distribution

• The bell-shaped curve

• Commonly referred to as the “law of error”

\[
p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right]
\]

• \( E(X) = \mu \), and \( \text{var}(X) = \sigma^2 \)
A normal family

• We refer to $\mu$ and $\sigma^2$ as parameters

• By specifying different values of the parameters, we obtain different members of the normal family
Recall...

- For any random variable $X$, the mean of $aX+b$ is $b$ and the variance is $a^2$.

- If $X$ has a standard normal distribution, we can construct other members of the family with $aX+b$. 
The *standard* normal distribution has mean 0 and variance 1.
X has a standard normal distribution
$3X \ (or \ \sigma^2 = 9)$
$5X \ (or \ \sigma^2 = 25)$
$3X \text{ (or } \sigma^2 = 9\text{)}$
$3X + 2$ (or $\sigma^2 = 9$ and $\mu = 2$)
$3X - 5$ (or $\sigma^2 = 9$ and $\mu = -5$)
Standard units

• Typically, we let $Z$ denote a random variable with the standard normal distribution.

• If $Z$ has a standard normal distribution, then $aZ+b$ again has a normal distribution but with mean $b$ and variance $a^2$.

• Similarly, if $X$ is normal with mean $\mu$ and variance $\sigma^2$, then $\frac{X-\mu}{\sigma}$ has a standard normal distribution.
Calculating probabilities

- Last time we saw that we computed the probability of events from a continuous density by calculating areas under the curve

- For the standard normal distribution, we can either look up values in a table or use the computer
<table>
<thead>
<tr>
<th>$z$</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
<td>0.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
<td>0.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
<td>0.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
<td>0.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
<td>0.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
<td>0.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
<td>0.7517</td>
<td>0.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
<td>0.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
<td>0.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8461</td>
<td>0.8485</td>
<td>0.8508</td>
<td>0.8531</td>
<td>0.8554</td>
<td>0.8577</td>
<td>0.8599</td>
<td>0.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8665</td>
<td>0.8686</td>
<td>0.8708</td>
<td>0.8729</td>
<td>0.8749</td>
<td>0.8770</td>
<td>0.8790</td>
<td>0.8810</td>
<td>0.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8907</td>
<td>0.8925</td>
<td>0.8944</td>
<td>0.8962</td>
<td>0.8980</td>
<td>0.8997</td>
<td>0.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9066</td>
<td>0.9082</td>
<td>0.9099</td>
<td>0.9115</td>
<td>0.9131</td>
<td>0.9147</td>
<td>0.9162</td>
<td>0.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9251</td>
<td>0.9265</td>
<td>0.9279</td>
<td>0.9292</td>
<td>0.9306</td>
<td>0.9319</td>
</tr>
</tbody>
</table>

The Standard Normal Table displays the cumulative probability $P(Z \leq z)$ for various values of $z$. This table is useful in statistics for determining probabilities related to the normal distribution.
Calculating probabilities

• Or, you can use a computing package like R

\[ \text{pnorm}(z, m=0, sd=1) \]
For other members of the family...

\[
\Pr( X \leq x ) = \Pr \left( \frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \right) = \Pr \left( Z \leq \frac{x - \mu}{\sigma} \right)
\]

where \( Z \) has a standard normal distribution
For other members of the family...

- So if $X$ is normal with mean 6 and variance 4, the chance that $X$ is at most 5 is

$$\Pr(X \leq 5) = \Pr\left(\frac{X - 6}{2} \leq \frac{5 - 6}{2}\right) = \Pr\left(Z \leq -\frac{1}{2}\right) = 0.3085$$
Calculating probabilities

• The chance that Z is greater than 5 is $1 - 0.3085 = 0.6915$

• By symmetry, the chance that Z is greater than 7 is also 0.3085 and the chance that it is less than 7 is 0.6915

• That means the chance that Z is between 5 and 7 is $0.6915 - 0.3085 = 0.3830$
Calculating probabilities

• The chance that Z is greater than 5 is $1-0.3085 = 0.6915$

• By symmetry, the chance that Z is greater than 7 is also 0.3085 and the chance that it is less than 7 is 0.6915

• That means the chance that Z is between 5 and 7 is $0.6915-0.3085 = 0.3830$
68%-95% rule

- For any normal random variable

\[ \text{pr}( \mu - \sigma \leq X \leq \mu + \sigma ) = 0.68 \]

and

\[ \text{pr}( \mu - 2\sigma \leq X \leq \mu + 2\sigma ) = 0.95 \]
Recall...

• For any pair of random variables $X_1$ and $X_2$, the mean of $X_1 + X_2$ is $\mu_1 + \mu_2$.

• If $X_1$ and $X_2$ are independent, then the variance of $X_1 + X_2$ is $\sigma_1^2 + \sigma_2^2$. 

\[
\mu_1 + \mu_2 = \mu_1 + \mu_2
\]

\[
\sigma_1^2 + \sigma_2^2 = \sigma_1^2 + \sigma_2^2
\]
For normals...

• If $X_1$ and $X_2$ are independent and normally distributed, then their sum $X_1 + X_2$ is also normal, with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$
2 normals: $\mu_1 = 3, \sigma^2_1 = 9$ and $\mu_2 = 7, \sigma^2_2 = 4$
Their sum is normal $\mu = \mu_1 + \mu_2 = 10$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2 = 13$
\( \mu_1 = -2, \mu_2 = 2 \) and \( \sigma^2_1 = \sigma^2_2 = 4 \)
This is NOT always the case

• For the normal family, the density curve for a sum of independent normals is again bell shaped

• In general, the form of the density curve corresponding to the sum of two random variables has to be computed (calculus)
Parameters and estimates

- A *parameter* is a numerical characteristic of a distribution

- Binomial: The number of trials $n$ and the success probability $p$ are parameters

- Normal: The mean and variance are parameters
Parameters and estimates

- We use data to produce an estimate of an unknown parameter
- Data often consist of a sample drawn from a probability distribution
Parameters and estimates

- Consider a normal distribution with mean $\mu$ and variance $\sigma^2$

- We let $X_1, X_2, \ldots, X_n$ denote a sample of size $n$ from this distribution
Male heights from the BRFSS data,
Sample mean = 70.16
Sample SD = 3.07
Parameters and estimates

• The sample mean is an estimate of \( \mu \)

\[
\bar{X} = \frac{1}{n} \sum_{i} X_i
\]
Parameters and estimates

• Each time we take a new sample, we get a different sample mean due to sampling variation

71 66 74 69 67 70 71 72 66 62 70

or

64 65 70 66 69 67 67 71 68 77 67
Parameters and estimates

- The expected value of the sample mean

\[ E\bar{X} = \frac{1}{n} \sum_{i} E X_i = \frac{1}{n} \sum_{i} \mu = \mu \]
Parameters and estimates

• The variance of the sample mean

$$\text{var} \bar{X} = \frac{1}{n^2} \sum_i \text{var} X_i = \frac{1}{n^2} \sum_i \sigma^2 = \frac{\sigma^2}{n}$$