R TUTORIAL, #10:
BINOMIAL DISTRIBUTIONS

The (>) symbol indicates something that you will type in.
A bullet (•) indicates what the R program should output (and other comments).

BINOMIAL COEFFICIENTS, PASCAL'S TRIANGLE, and LOOPS

• Find $\binom{5}{2}$, or ${}_5C_2$.
> Type: choose(5,2)

• Give all binomial coefficients of the form $\binom{5}{x}$.
> Type: choose(5, 0:5)

• Use a loop to print the first several rows of Pascal's triangle.
> Type: for (n in 0:10) print(choose(n, 0:n))

FIVE DICE EXAMPLE (COUNTING '4's)

• Find the probability of getting two '4's among five dice.
> Type: dbinom(2, size=5, prob=1/6)
    • 'd' stands for "density," as in "probability density function."

• We will find $P(2)$ by using the binomial probability formula.
> Type: choose(5,2) * (1/6)^2 * (5/6)^3
    • $P(2)=\binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$

• We will obtain the table for $\text{Bin}\left(n=5, p=\frac{1}{6}\right)$.
> Type: probs = dbinom(0:5, size=5, prob=1/6)
> Type: data.frame(0:5, probs)

• Verify that the sum of the probabilities is 1.
> Type: sum(probs)

BINOMIAL PROBABILITY SPIKE PLOTS

• We can do a spike plot for the distribution $\text{Bin}\left(n = 5, p = \frac{1}{6}\right)$.

> Type: plot(0:5, probs, type="h", xlim=c(0,5), ylim=c(0,.5))
> Type: points(0:5, probs, pch=16, cex=2)

FAIR COIN EXAMPLE (COUNT HEADS IN 100 FLIPS)

• We will obtain the table for $\text{Bin}\left(n = 100, p = \frac{1}{2}\right)$.

> Type: probs2 = dbinom(0:100, size=100, prob=1/2)
> Type: data.frame(0:100, probs2)

- We get probabilities using scientific notation.
- Observe that: $P(0) \approx 7.9 \times 10^{-31}$

> Type: round(data.frame(0:100, probs2), digits=5)

- We now get the probabilities to five decimal places.

• Verify that the sum of the probabilities is 1.
> Type: sum(probs2)

• We can do a spike plot for the distribution.
> Type: plot(0:100, probs2, type="h", xlim=c(0,100), ylim=c(0,.1))
> Type: points(0:100, probs2, pch=16, cex=.5)

• Find the probability of getting at least 59 heads.
> Type: sum(dbinom(59:100, size=100, prob=1/2))

- You could also do: sum(probs2[60:101])
- Remember that probs2[1] corresponds to $P(0)$, not $P(1)$.

EXPERIMENTING WITH BINOMIAL DISTRIBUTIONS;
MONTY HALL REVISITED;
MEAN, VARIANCE, and STANDARD DEVIATION

• Let's start with the "Monty Hall" distribution $\text{Bin}\left(n = 18, p = \frac{2}{3}\right)$.

• Then, you can alter the parameters *n* and *p* and see what happens.
> Type: n = 18
> Type: p = 2/3
> Type: probs3 = dbinom(0:n, size=n, prob=p)
> Type: round(data.frame(0:n, probs3), digits=5)

• We can do a spike plot for the distribution.
> Type: plot(0:n, probs3, type="h", xlim=c(0,n), ylim=c(0, max(probs3)))
> Type: points(0:n, probs3, pch=16, cex=1)

• Let's find the mean (or expected value) of this distribution.
> Type: n*p

- For a binomial $\text{Bin}(n, p)$ distribution, $\mu = np$.

• Let's find the variance of this distribution.
> Type: n*p*(1-p)

- For a binomial $\text{Bin}(n, p)$ distribution, $\sigma^2 = npq = np(1-p)$.

• Let's find the standard deviation of this distribution.
> Type: sqrt(n*p*(1-p))

- For a binomial $\text{Bin}(n, p)$ distribution, $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$.

COMPARING SAMPLE RESULTS WITH THEORETICAL RESULTS: MORE MONTY HALL

• Let's go back to our Monty Hall distribution.
> Type: probs3 = dbinom(0:18, size=18, prob=2/3)

• Let's get a sample (of size 1) from the Monty Hall distribution.
> Type: rbinom(1, size=18, p=2/3)

• Let's get a sample (of size 1000) from the Monty Hall distribution.
> Type: results = rbinom(1000, size=18, p=2/3)
> Type: results

> Type: mean(results)
- Compare this sample mean with the theoretical mean of the distribution we found earlier.

> Type: var(results)
- We do want the sample variance here.
- Compare this sample variance with the theoretical variance of the distribution we found earlier.

> Type: sd(results)
- Compare this sample standard deviation with the theoretical standard deviation of the distribution we found earlier.

> Type: hist(results, prob=T, breaks=seq(-0.5, 18.5, by=1), ylim=c(0, max(probs3)))
- We see a relative frequency histogram of our results.
- For example, the interval from 3.5 to 4.5 corresponds to "4."

• Compare this with the spike plot of the theoretical distribution we created earlier.
> (Open a new "Quartz Device Window.")
> Type: plot(0:n, probs3, type="h", xlim=c(0,n), ylim=c(0, max(probs3)))
> Type: points(0:n, probs3, pch=16, cex=1)

GO BACK TO THE PREVIOUS SECTION, AND EXPERIMENT!