Chapter 15 - Sampling Distribution Models

15.1 Sampling Distribution of a Proportion

Back in Chapter 10 we learned about **Sampling Variability** (The natural occurrence of the sample statistic to vary from sample to sample).

The **Sampling Distribution Model** shows the behavior of the sample statistic over all possible samples of the same size n.

Yesterday we simulated a Sampling Distribution Model

- The model was unimodal and symmetric (like a Normal curve)
- The model was centered around the population mean, μ.

Sampling Distribution Model for a Proportion

If assumptions and conditions are met, then the sampling distribution of proportion is modeled by a Normal model with
mean equal to the true proportion, $\mu_{\hat{p}} = p$, and

standard deviation equal to $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

- i.e. $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

$\frac{\mu}{n} = \frac{np}{n} = p$

$\frac{\sigma}{n} = \frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$

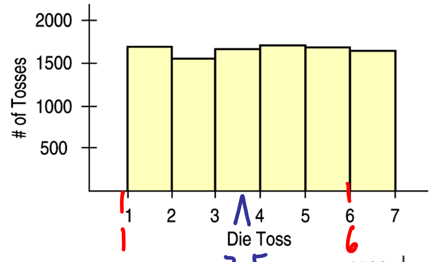
15.2 When does the Normal Model Work? Assumptions and Conditions

In order to find a model for the distribution of the Sample Proportion, certain assumptions and conditions must be satisfied

- **Independence Assumption**: The sampled values must be independent of each other.
- **Randomization Condition**: Data should come from randomized source.
- **10% Condition**: The sample size, n, must be no larger than 10% of the population. If the sample size is larger than 10% of the population the remaining individuals are no longer independent of each other.
- **Success/Failure Condition**: The sample size has to be big enough so that $np \geq 10$ and $nq \geq 10$ - i.e., there needs to be at least 10 successes and 10 failures.

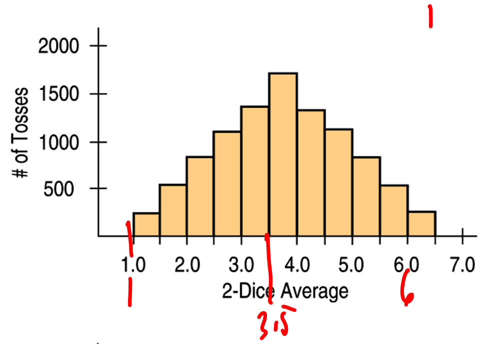
15.3 The Sampling Distribution of Other Statistics

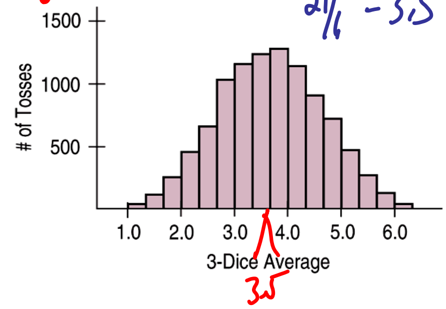
uniform
symmetric

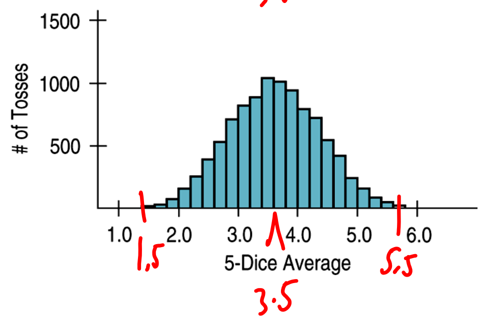


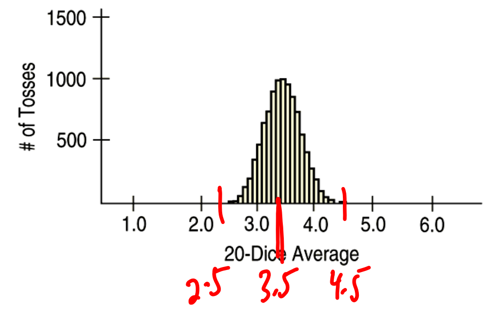
X	$P(x)$	$x \cdot P(x)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6

$\frac{21}{6} = 3.5 = \mu$









15.4 The Central Limit Theorem: The Fundamental Theorem of Statistics

Sampling Distribution Model for a Mean

For the mean, the only assumption needed is that the observations must be independent and random. Also want the sample size, *n*, to be no more than 10% of the population.

If the sample size is large enough then the population distribution does not matter.

Central Limit Theorem (The Fundamental Theorem of Statistics)

The mean of a random sample has a sampling distribution that is approximately normal with mean, $\mu_{\overline{x}} = \mu$, and

standard deviation, $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}$.

- i.e. $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

The larger the sample, the better the approximation.

Since population parameters are rarely known estimates called **Standard Error** will be used for the standard deviation:

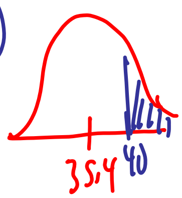
- For the sample proportion: $SE_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$
- For the sample mean: $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$

p.425

(50)

$\mu = 35.4$ inches $\quad \sigma = 4.2$ inches

a)



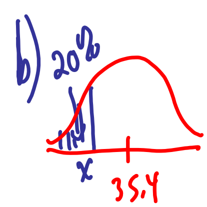
Normalcdf(40, 100000, 35.4, 4.2)

left endpt, right endpt, μ, σ

$P(X > 40 \text{ inches}) \approx .1367$

13.7% of years will have rainfall over 40 inches.

b)



InvNorm(.2, 35.4, 4.2) $\quad x = 31.9$

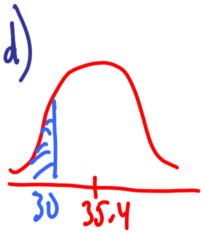
% to the left, μ, σ

Driest 20% of years have rainfall below 31.9 inches.

c) Years should be independent, n=4 is less than 10% of all years (10% cond.)
Random Not mentioned. Central Limit Theorem Says Distribution is Normal with

$\mu_{\bar{x}} = \mu = 35.4$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4.2}{\sqrt{4}} = 2.1 \quad N(35.4, 2.1)$

d)



Normalcdf(-100000, 30, 35.4, 2.1)

$P(\bar{x} < 30'') \approx .005$

p.422

34 $p = .92 \quad q = 1 - .92 = .08 \quad n = 160$

Seeds should be independent. seeds should be random

10% cond: $n = 160$ is less than 10% of all this type of seed.

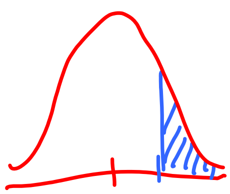
Success/Failure condition: $np = 160 \times .92 = 147.2 \geq 10$ Conditions are met

$nq = 160 \times .08 = 12.8 \geq 10$

Distribution is Approximately Normal with

$\mu_{\hat{p}} = p = .92$ & $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.92 \times .08}{160}} \approx .0214$

Normalcdf(.95, 100000, .92, .0214)



.92 .95

$P(\hat{p} > 95\%) \approx .0805$

8.05% chance of this packet of seeds germinating at a rate higher than 95%.

p.423

36 $p = .04 \quad q = 1 - .04 = .96 \quad n = 732$

Newborns should be independent, Random?

10% cond: $n = 732$ is less than 10% of all Newborns

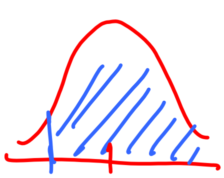
S/F cond: $np = 732 \times .04 = 29.28 \geq 10$ Conditions are met.

$nq = 732 \times .96 = 702.72 \geq 10$

Distribution is approximately Normal with

$\mu_{\hat{p}} = p = .04$ and $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.04 \times .96}{732}} \approx .0072$

$\frac{20}{732} = .0273$ at least 20 out of the 732 or $\hat{p} \geq \frac{20}{732}$



.0273 .04

Normalcdf(.0273, 100000, .04, .0072)

$P(\hat{p} \geq .0273) \approx .9611$

96.1% chance that at least 20 Newborns out of the 732 will have the gene for juvenile diabetes.