Complete on your own for Bonus Note points

Example 10a: What is the probability of drawing 3 cards from a deck *with replacement* and getting no hearts?

$\left(\frac{39}{52}\right)^3$

Example 10b: What is the probability of drawing 3 cards from a deck *without replacement* and getting no hearts?

$\frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$

Example 11a: What is the probability of drawing 3 cards from a deck *with replacement* and getting at least one heart?

$P(\text{At least one } \heartsuit) = 1 - P(\text{No } \heartsuit) = 1 - \left(\frac{39}{52}\right)^3$

Example 11b: What is the probability of drawing 3 cards from a deck *without replacement* and getting at least one heart?

$1 - \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$

Example 12a: What is the probability of drawing 3 cards from a deck *with replacement* and getting all hearts?

$P(\text{All } \heartsuit) = \left(\frac{13}{52}\right)^3$

Example 12b: What is the probability of drawing 3 cards from a deck *without replacement* and getting all hearts?

$\frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}$

Example 13a: What is the probability of drawing 3 cards from a deck *with replacement* and not getting all hearts?

$P(\text{Not all } \heartsuit) = 1 - P(\text{All } \heartsuit) = 1 - \left(\frac{13}{52}\right)^3$

Example 13b: What is the probability of drawing 3 cards from a deck *without replacement* and not getting all hearts?

$1 - \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}$

Example 4:

36 draws are made at random with replacement from the box containing 4 tickets: $\boxed{0}\boxed{2}\boxed{2}\boxed{8}$

avg = 3   SD = 3

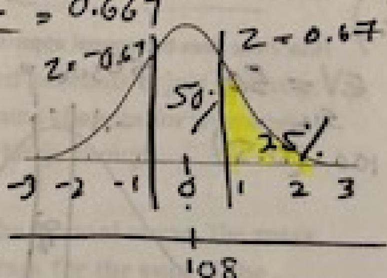
a) EV_{sum} = 108 SE_{sum} = 18

$EV_{sum} = n \cdot avg = 36 \cdot 3 = 108$

$SE_{sum} = \text{SD of box} \times \sqrt{n} = 3 \times \sqrt{36} = 18$

b) What is the probability that the sum of the 36 draws would be greater than 120?

$Z = \dfrac{Val - EV_{sum}}{SE_{sum}} = \dfrac{120-108}{18} = \dfrac{12}{18} = 0.667$



Do on your own

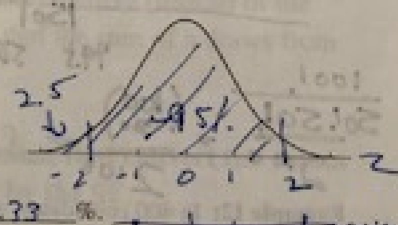
c) EV_{avg} = 3 SE_{avg} = 0.5

EV_{avg} = avg of box $SE_{avg} = \dfrac{SD}{\sqrt{n}} = \dfrac{3}{\sqrt{36}} = 0.5$

d) What is the probability that the average of the 36 draws would be less than 4?

$Z = \dfrac{val - EV_{avg}}{SE_{avg}} = \dfrac{4-3}{0.5} = 2$

middle area + left tail
95% + 2.5 = $\boxed{97.5\%}$



e) EV of the percent of $\boxed{2}$'s in 36 draws = 50% with SE 8.33 %.

$EV_{\%}$ = % in box

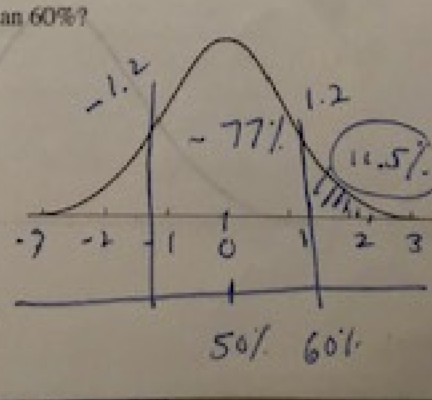
To find the SE of the percent $\boxed{2}$'s first change $\boxed{0}\boxed{2}\boxed{2}\boxed{8}$ >>>>> to a 0-1 box: $\boxed{0\ 1\ 1\ 0}$

$SE_{\%} = \dfrac{SD}{\sqrt{n}} \times 100 = \dfrac{0.5}{\sqrt{36}} \times 100 = 8.33\%$

f) What is the probability that the percent of 2's would be more than 60%?

$Z = \dfrac{Val - EV_{\%}}{SE_{\%}} = \dfrac{60-50}{8.33} = 1.2$

Right tail $= \dfrac{100-77}{2}$ $\boxed{11.5\%}$



Z

%

Fill in for Bonus Points

One-tail vs Two-tail z-tests

Specifying a direction to the alternative hypothesis is called a "one-tailed z-test". Not specifying a direction to the alternative, just specifying an inequality is called "a two-tailed z-test".

Both tests yield the same Z value, but the p-value of the 2-tailed test is double the p-value of the one-tailed test because it includes the area of *both* tails instead of one tail.

A chi-square Independence test with 2 variables and 2 categories is equivalent to a 2-tailed z-test. *

Which test should be used, one-tailed or two-tailed? That depends on which alternative makes more sense for a given situation. But since we can translate back and forth simply by doubling (or halving) the p-value it doesn't really matter which one we use as long as you clearly state which alternative you used.

Here's the "love at first sight" example on p.80 as both a one-tailed and two-tailed test. See p.80 same p-values as p.88

$H_0: p_M = p_F$

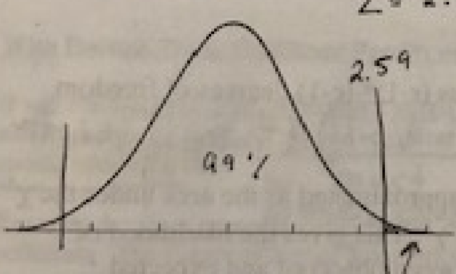
One Tail

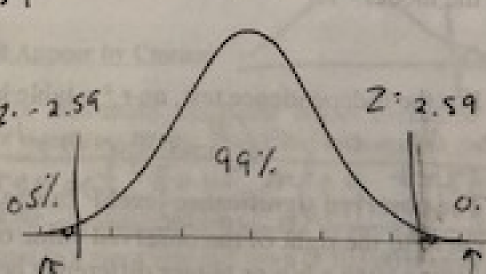
$H_a: p_M > p_F$

Two Tail

$H_a: p_M \neq p_F$

Z = 2.59





$p = \frac{100\% - 99\%}{2} = 0.5\%$ Area of 1 tail

$p = 100 - 99\% = 1\%$ Area of 2 tails

$\chi^2 = 6.67$ df = 1 p-value = 1% (p.88) same as 2-sided Z (p.80)

Chapter 17: Limitations of Significance Tests

Fill in for Bonus pts.

1 Statistical Significance vs. Practical Significance

Significance tests can only tell you whether or not a difference is likely due to chance—not whether a difference is important.

Example 1: Suppose investigators want to compare IQ scores of men and women. They take a simple random sample of 10,000 men and an independent simple random sample of 10,000 women. The men average 100 on the IQ test and their SD is 20, while the women average 101 with an SD of 20. The investigators do a 2-sample z-test and conclude that the difference is highly statistically significant.

However, even though the 1-point difference is *statistically* significant, it's basically meaningless. For all practical purposes, the difference is insignificant, no matter how statistically significant it is.

* *Note that with a large enough sample, any difference, no matter how tiny, can be found statistically significant.* (We can make our standard errors arbitrarily small by increasing the sample size.) That doesn't make the tiny difference important.

* *And if our samples are too small, very important differences may not show up to be statistically significant because the statistics are weak.*

2 With Enough Tests, Significant Results will Appear by Chance

P-hacking, Data Snooping All cheating!

If you run 100 tests, you can expect 5 to result in "statistically significant" results, even if the null hypothesis is true. So beware of claims of significant results when many tests are run (or should have been run). Report how many tests were run. You can't just "cherry-pick" the significant results

Example 2: Liver cancer is a rare disease, which may sometimes be caused by environmental pollutants. The chance of having 2 or more cases in a given year in a town of 10,000 is small—about 0.5%. A cluster of liver cancers (several in a small community) prompts a search for causes, like a contaminated water supply.

Suppose an environmental protection agency reviews the incidence of liver cancer in 100 towns of this size over a 10-year period looking for possible environmental contaminants. How many clusters would you expect them to find just by chance, even if none of the towns were contaminated?

100 towns * 10 years = 1,000 town/yr com.
So we'd expect 0.5% of 1000 = $\boxed{5}$ clusters even when all the towns are clean!

Using the exact probability distribution to compute p-values.
For small sample sizes, the normal curve is not such a good approximation of the probability distribution of the rank sum or U statistics. But that's not a problem because once you convert data to ranks or U counts, the numbers are simple enough that it's not hard to calculating exact probability distributions for small samples by hand.

More conveniently, there are tables available for the exact probability distribution of the U stat (but not the rank sum stat), so most people use U tables to find critical values.

The table below gives the critical value of U at $p \leq 0.05$ for the 2-tail test (or $p \leq 0.025$ for the 1-tail test). N_1 and N_2 are the sample sizes of the smaller and larger groups, respectively. These are the critical values for the *smaller U stat*, so your observed U (the smaller one) has to be ≤ the critical U listed to reject the null.

Compare our U stat in the last example to the critical value we would have needed to reject the null?

U^* at $\alpha = 0.05$ for smaller U-stat.

N_2 (size of larger group)

N_1 sample size of smaller group

	5	6	7	8	9	10	11	12	13	14	15
5	2	3	5	6	7	8	9	11	12	13	14
6		5	6	8	10	11	13	14	16	17	19
7			8	$\boxed{10}$	12	14	16	18	20	22	24
8				13	15	17	19	22	24	26	29
9					17	20	23	26	28	31	34
10						23	26	29	33	36	39
11							30	33	37	40	44
12								37	41	45	49
13									45	50	54
14										55	59
15											64

Is our smaller U stat = 27.5 ≤ U^* stat = 10 ?
No, so we cannot reject null. (same as bigger U stat ≥ 46.)

f) Rearrange the exam scores to find the U stat closest to the critical value listed without going over.

First find smallest possible U-stat (see below).

For bonus note back credit

	Group A	Group B:
Scores:	11, 70, 75, 85, 88, 92, 96	60, 65, 70, 79, 81, 90, 95, 100
Rearranged Scores	60, 65, 70, 70, 75, 79, 90	11, 81, 85, 88, 92, 95, 96, 100
U counts:	1+1+1+1+1+1+4 = 10	0+6+6+6+7+7+7+7 = 46

smallest possible U stat, then switch 11 ⟷ 90

A: 11, 60, 65, 70, 70, 75, 79
U: 0+0+0+0+0+0+0 = 0

B: 81, 85, 88, 90, 92, 95, 96, 100
7 7 7 7 7 7 7 7 = 56