What are odds?
The odds of outcome 1 versus outcome 2 are the probability (or frequency) of outcome 1 divided by the probability (or frequency) of outcome 2.

Question 1: Odds Example

There are 5 pink marbles, 2 blue marbles, and 8 purple marbles.

- What is the probability of picking a blue marble? (Answer: 2/15).
- What are the odds in favor of picking a blue marble? (Answer: 2/13).

Question 2: Odds Ratio Example

Suppose that seven out of 10 males are admitted to an engineering school while three of 10 females are admitted.

- The probabilities for admitting a male are, p = 7/10 = 0.7 (q = 1 - 0.7 = 0.3)
- Here are the same probabilities for females, p = 3/10 = 0.3 (q = 1 - 0.3 = 0.7)

Now we can use the probabilities to compute the admission odds for both males and females,

- *odds(male)* = 0.7/0.3 = 2.33333
- *odds(female)* = 0.3/0.7 = 0.42857

Next, we compute the odds ratio for admission,

$$OR = 2.3333/0.42857 = 5.44$$

Thus, for a male, the odds of being admitted are 5.44 times as large than the odds for a female being admitted.

Question 3: Odds Ratio Example

- Suppose that in a sample of 100 men, 90 drank wine in the previous week, while in a sample of 100 women only 20 drank wine in the same period.
- The odds of a man drinking wine are 90 to 10, or 9:1, while the odds of a woman drinking wine are only 20 to 80, or 1:4 = 0.25:1.
- The odds ratio is thus 9/0.25, or 36, showing that men are much more likely to drink wine than women.

- The detailed calculation is:

$$\frac{0.9/0.1}{0.2/0.8} = \frac{0.9 \times 0.8}{0.1 \times 0.2} = \frac{0.72}{0.02} = 36$$

- This example also shows how odds ratios are sometimes sensitive in stating relative positions: in this sample men are $90/20 = 4.5$ times more likely to have drunk wine than women, but have 36 times the odds.
- The logarithm of the odds ratio, the difference of the logits of the probabilities, tempers this effect, and also makes the measure symmetric with respect to the ordering of groups.
- For example, using natural logarithms, an odds ratio of 36/1 maps to 3.584, and an odds ratio of 1/36 maps to -3.584.

Question 4: Odds Ratio Example

These data are taken from the British Election Study 2005 pre-campaign and post-election panel data. We will consider the propensity to vote (sometimes called turnout) as the dependent variable, which has 2 categories. out to vote.

gender of respondent * vote2005 Crosstabulation

			vote2005		
			didn't vote	voted	Total
gender of respondent	male	Count	491	1346	1837
		% within gender of respondent	26.7%	73.3%	100.0%
		% within vote2005	45.5%	43.8%	44.2%
	female	Count	587	1729	2316
		% within gender of respondent	25.3%	74.7%	100.0%
		% within vote2005	54.5%	56.2%	55.8%
Total		Count	1078	3075	4153
		% within gender of respondent	26.0%	74.0%	100.0%
		% within vote2005	100.0%	100.0%	100.0%

- The odds of a male turning out to vote are:

$$1346/491 = 2.741$$

- The odds of female turning out to vote are

$$1729/587 = 2.945$$

- The Odds ratio (female: male) are

$$\frac{1729/587}{1346/491} = \frac{2.945}{2.741} = 1.074$$

The Logit

- The convention for binomial logistic regression is to code the dependent class of greatest interest as 1 and the other class as 0, because the coding will affect the odds ratios and slope estimates.
- The logit(p) is the log (to base e) of the odds ratio or likelihood ratio that the dependent variable is 1. In symbols it is defined as:

$$logit(p) = ln\left(\frac{p}{(1-p)}\right)$$

- Whereas p can only range from 0 to 1, logit(p) scale ranges from negative infinity to positive infinity and is symmetrical around the logit of 0.5 (which is zero)

Question 5: Logistic Transformation

- Given that $p_i = 0.2$, compute η_i.

$$\eta_i = \log\left(\frac{0.2}{1-0.2}\right) = \log\left(\frac{0.2}{0.8}\right)$$

$$\eta_i = \log(0.25) = -1.386$$

- Given that $\eta_i = 2.3$, compute p_i.

$$\pi_i = \frac{e^{2.3}}{1+e^{2.3}} = \frac{9.974}{1+9.974} = 0.908$$

Question 6: Classification

Let us suppose that the probability of survival of a marine species of fauna is dependent on pollution, depth and water temperature. Suppose the logit for the logistic regression was computed as follows:

$$\eta_i = 0.14 + 0.76x_1 - 0.093x_2 + 1.2x_3$$

Variables	case 1	case 2
Pollution(x_1)	6.0	1.9
Depth (x_2)	51	99
Temp (x_3)	3.0	2.9

Compute the probability of success for both case 1 and case 2.

- case 1 $\eta_1 = 0.14 + (0.76 \times 6) - (0.093 \times 51) + (1.2 \times 3) = 3.557$
- case 2 $\eta_2 = 0.14 + (0.76 \times 1.9) - (0.093 \times 99) + (1.2 \times 2.9) = -4.143$

The probabilities for success are therefore:

$$\pi_1 = \frac{e^{3.557}}{1 + e^{3.557}} = \frac{35.057}{1 + 35.057} = 0.972$$

$$\pi_2 = \frac{e^{-4.143}}{1 + e^{-4.143}} = \frac{0.0158}{1 + 0.0158} = 0.0156$$

Question 7: Wald Test

For the logistic regression model described in the SPSS output below, state the Odds Ratio for each variable. State the 95% confidence interval for the Odds Ratio in each case. Comment on this significance of each predictor variable.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	age	.085	.028	9.132	1	.003	1.089	1.030	1.151
	weight	.006	.022	.065	1	.799	1.006	.962	1.051
	gender(1)	1.950	.842	5.356	1	.021	7.026	1.348	36.625
	VO2max	-.099	.048	4.266	1	.039	.906	.824	.995
	Constant	-1.676	3.336	.253	1	.615	.187		

a. Variable(s) entered on step 1: age, weight, gender, VO2max.