6) Normal Distribution

Vitor Kamada

June 2018

Reference

Tables, Graphics, and Figures from

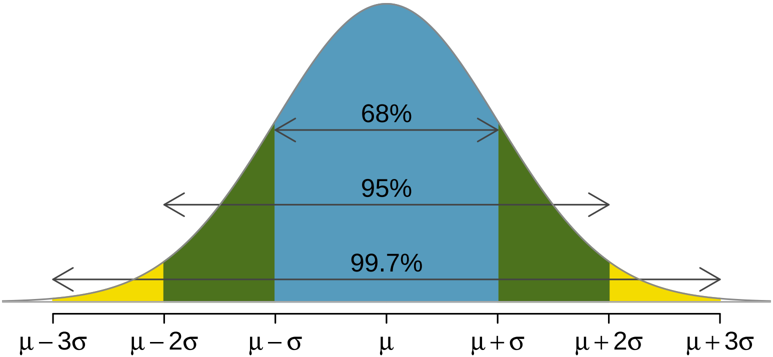
Introductory Statistics with Randomization and Simulation

Diez et al. (2014): Ch 2 - Foundation for inference

Central Limit Theorem (CLT)

$$Y = X_1 + X_2 + X_3 + ... + X_n$$

If Xs are independent and n large enough, then $Y \sim N(\mu, \sigma^2)$



Standardizing with Z scores

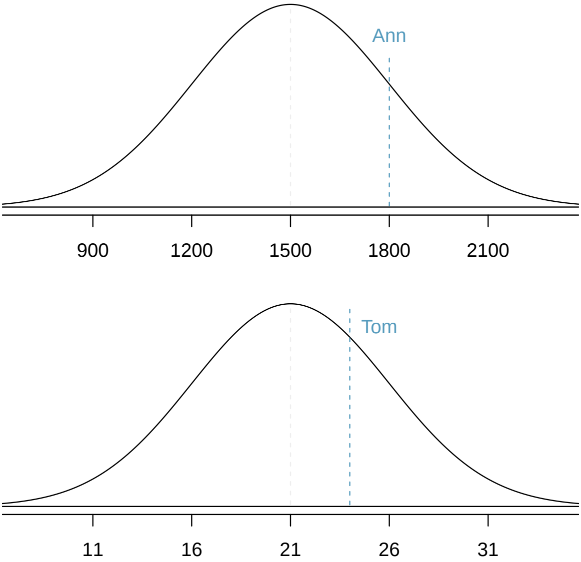
Ann scored 1800 on SAT

Tom scored 24 on his ACT

	SAT	ACT
Mean	1500	21
SD	300	5

$$Z = \frac{X - \mu}{\sigma} = \frac{24 - 21}{5} = 0.6$$

SAT and ACT



Normal Probability Table

Z	Second decimal place of Z									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

$$P(z < 0.43) \cong 66.6\%$$

$$P(z < 0.84) \cong 80\%$$

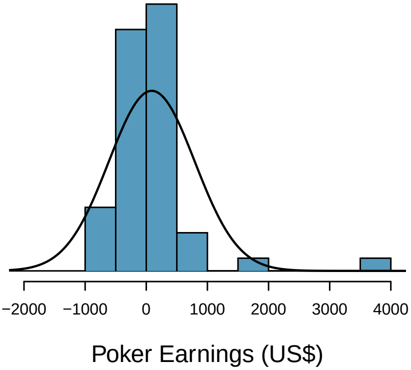
```
import scipy.stats
scipy.stats.norm(0, 1).cdf(0.43)
```

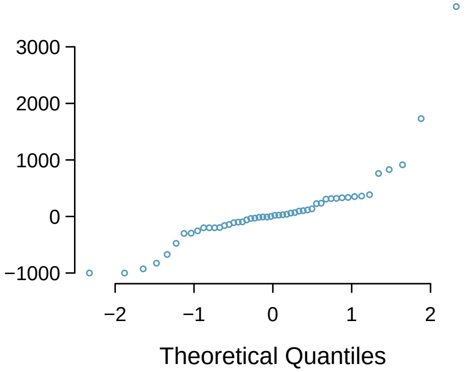
0.6664022

```
scipy.stats.norm(0, 1).ppf(0.6664)
```

0.43

Normal Probability Plot





Opportunity Cost

$$H_0 : p_t = p_c \text{ vs } H_A : p_t > p_c$$

	decision		
	buy DVD	not buy DVD	Total
control group	56	19	75
treatment group	41	34	75
Total	97	53	150

	decision		
	buy DVD	not buy DVD	Total
control group	0.747	0.253	1.000
treatment group	0.547	0.453	1.000
Total	0.647	0.353	1.000

Standard Error

$$\hat{p}_t = 0.453 \text{ and } \hat{p}_c = 0.253$$

$$n_t = 75 \text{ and } n_c = 75$$

$$SE_{p_t - p_c} = \sqrt{\frac{p_t(1-p_t)}{n_t} + \frac{p_c(1-p_c)}{n_c}} = 0.078$$

$$Z = \frac{p_t - p_c - 0}{SE_{p_t - p_c}} = \frac{0.2-0}{0.078} = 2.56$$

1-scipy.stats.norm(0, 1).cdf(2.56)

$$P(z > 2.56) = 0.0052$$

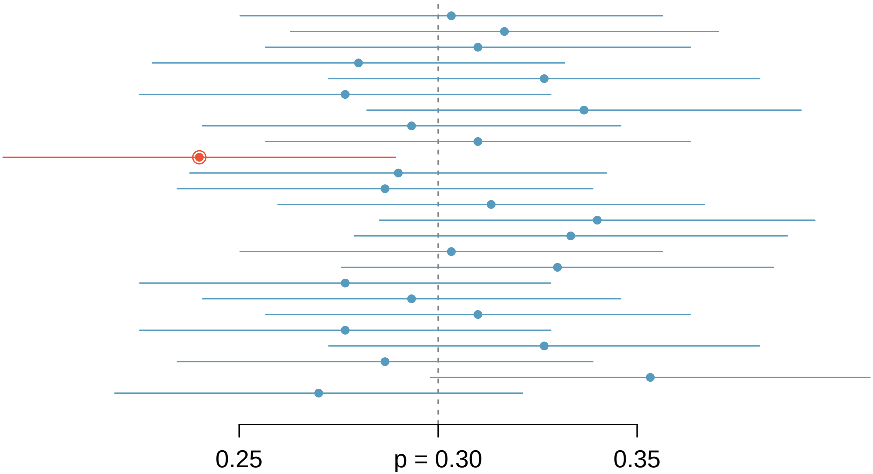
95% Confidence Interval

$$p \pm z_{2.5\%} \times SE$$

$$0.2 \pm 1.96 \times 0.078 = [0.047, 0.353]$$

We are 95% confident that the DVD purchase rate resulting from the treatment is between 4.7% and 35.3% lower than in the control group.

95% Confidence Intervals for $p = 0.3$



Margin of Error (ME)

$$p \pm z_{2.5\%} \times SE$$

$$0.2 \pm 1.96 \times 0.078 = [0.047, 0.353]$$

$$ME = z_{2.5\%} \times SE$$

$$= 1.96 \times 0.078 \cong 0.1528 = 15.28\%$$

Margin of Error and Sample Sizes

n	Margin of Error
100	10%
400	5%
625	4%
1,112	3%
2,500	2%
10,000	1%

Presidential Election in 1936

Companies	Sample Size	Landon	Roosevelt
Literary Digest	10M	57%	43%
Gallup	50K	44%	56%
Actual Result		37%	62%