CS626: Speech, NLP and Web

Hypothesis Testing

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

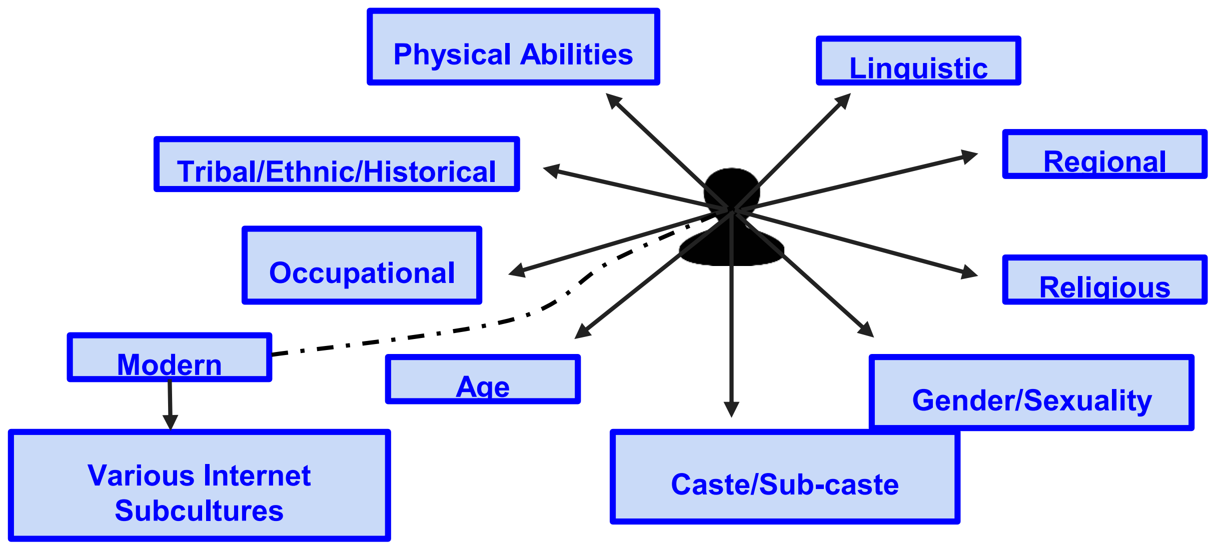
Week 11 of 14th October, 2024

1-slide recap of week of 7^{th} Oct

- Bias: < S, L, T, C, R >
- Stereotyping

Fact $\xrightarrow{\textit{Overgeneralized}}$ Stereotype $\xrightarrow{\textit{Opinionated}}$ Bias

- Bias types



- Role of probability: the bridge problem- deterministic (conservative, liberal); probabilistic

How to read the Z-score Table

A Practical Problem

- A bridge is being built. The weight it can tolerate has a distribution with $\mu=400$ and $\sigma=40$. A car that goes on the bridge has weight distribution given by $\mu=3$ and $\sigma=0.3$. We want the probability of damage to the bridge to be less than 0.1. How many cars can we allow to go on the bridge?

We want

What no. of cars will cause the damage probability to exceed 0.1?

Probability[(W_{total}-$W_{tolerance}$)>0] > 0.1

LHS is a function of N

W_{total} is normal by CLT, with $\mu=3N$ and $\sigma^2=0.09N$

Convert to Standard Normal Form

$$z \equiv \frac{(W_{total} - W_{tolerance}) - (3N - 400)}{\sqrt{0.09N + 1600}}$$

We want this event…

$$(W_{total} - W_{tolerance}) > 0$$

$$\Rightarrow \frac{(W_{total} - W_{tolerance}) - (3N - 400)}{\sqrt{0.09N + 1600}} > \frac{-(3N - 400)}{\sqrt{0.09N + 1600}}$$

$$\Rightarrow z > \frac{-(3N - 400)}{\sqrt{0.09N + 1600}}$$

When will this Probability exceed 0.1

$$P\left(z > \frac{-(3N-400)}{\sqrt{0.09N+1600}} \right) > 0.1$$

Solving this gives N <= 117

How?

V=1.29

Standard Normal Probabilities

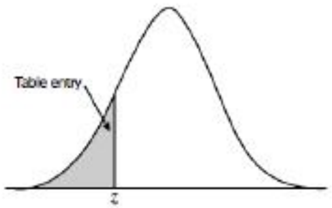


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard Normal Probabilities

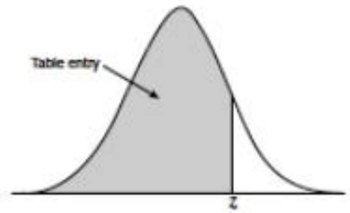


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Get N from…

$$1.29 = \frac{-(3N - 400)}{\sqrt{1600 + 0.09N}}$$

N= ~117

Use the rows and cols in z-score table

The value in any cell gives the area under the standard normal curve from **minus infinity to a value V of Z**

V is obtained by ADDING the column heading for the cell to the decimal part of the row heading of the cell

For example, the value of 0.9015 is obtained for the row:1.2 and col:0.09, i.e., for Z=1.29

Terminology for Test of hypothesis

Terminology

Null and Alternative Hypothesis

- H_0: Null Hypothesis → the hypothesis we want to reject with observation

- H_A or H_1: Alternative Hypothesis → opposite of H_0

- We use the sample statistics, trying to reject H_0

Type I and Type II error

- Type I: incorrectly reject H_0, when it should have been accepted.

- Type II: incorrectly accept H_0 when it should have been rejected.

More on H_0

- The data would be unlikely to occur if the null hypothesis were true.

- In logical form:

$$N_H \;\; |\text{-} \; {\sim}D$$

- Where N_H is the proposition “null hypothesis true” and D is the proposition “Data occurs”

Showing the methodology with a “coin toss” example

The setting

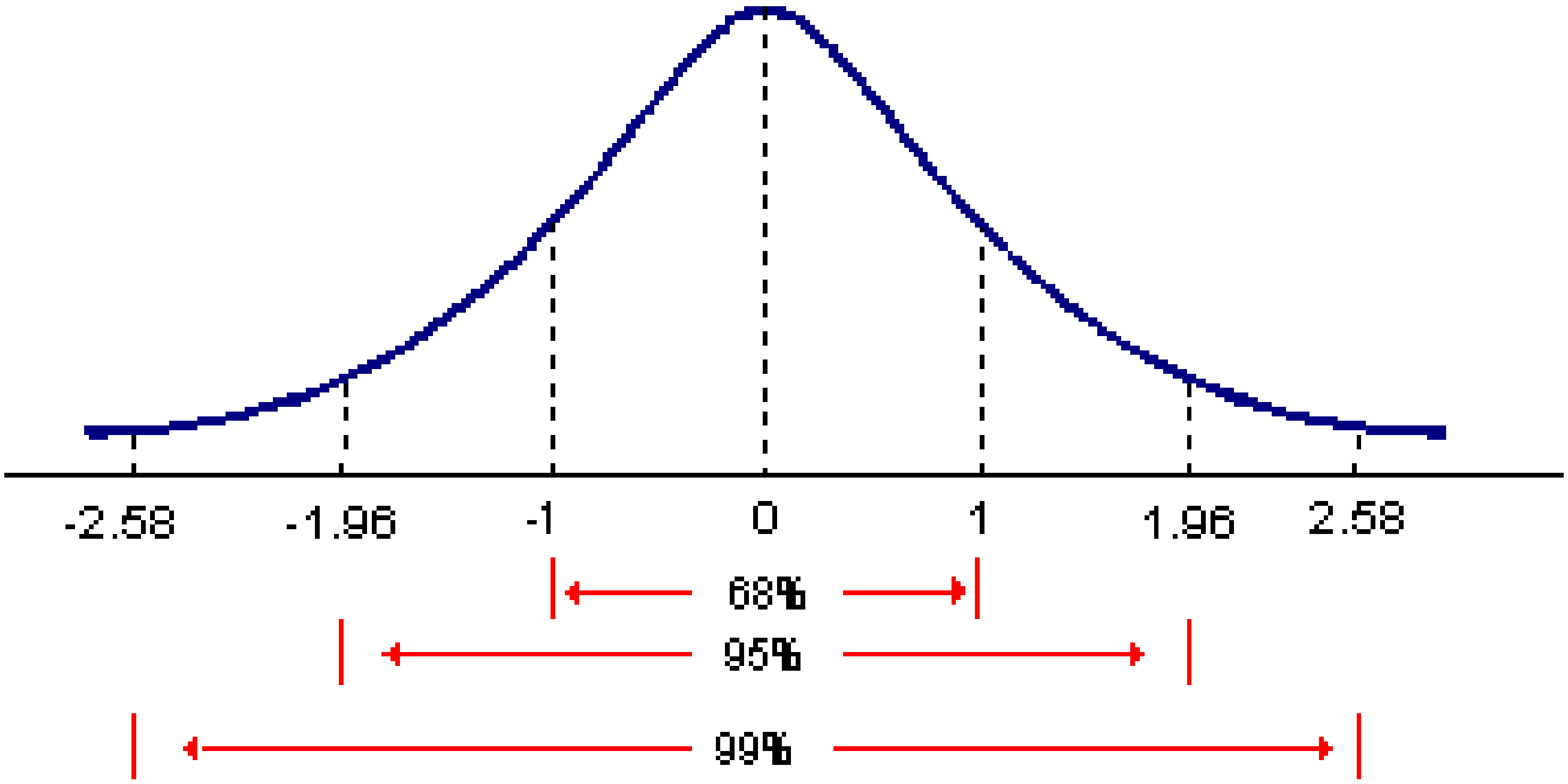
- A coin is tossed 100 time; there are 70 Heads
- H_0: coin is unbiased, i.e., p=0.5 (p is the parameter= probability of Head)
- Obviously, the null hypothesis is false
- How to show this?

Use normal approximation

- The approximating normal distribution has *mean=Np=100 X 0.5= 50 (>5)*
- *Standard deviation= sqrt[Np(1-p)]=sqrt(100 X 0.5 X 0.5)=sqrt(25)=5*
- *Z=(X-50)/5*

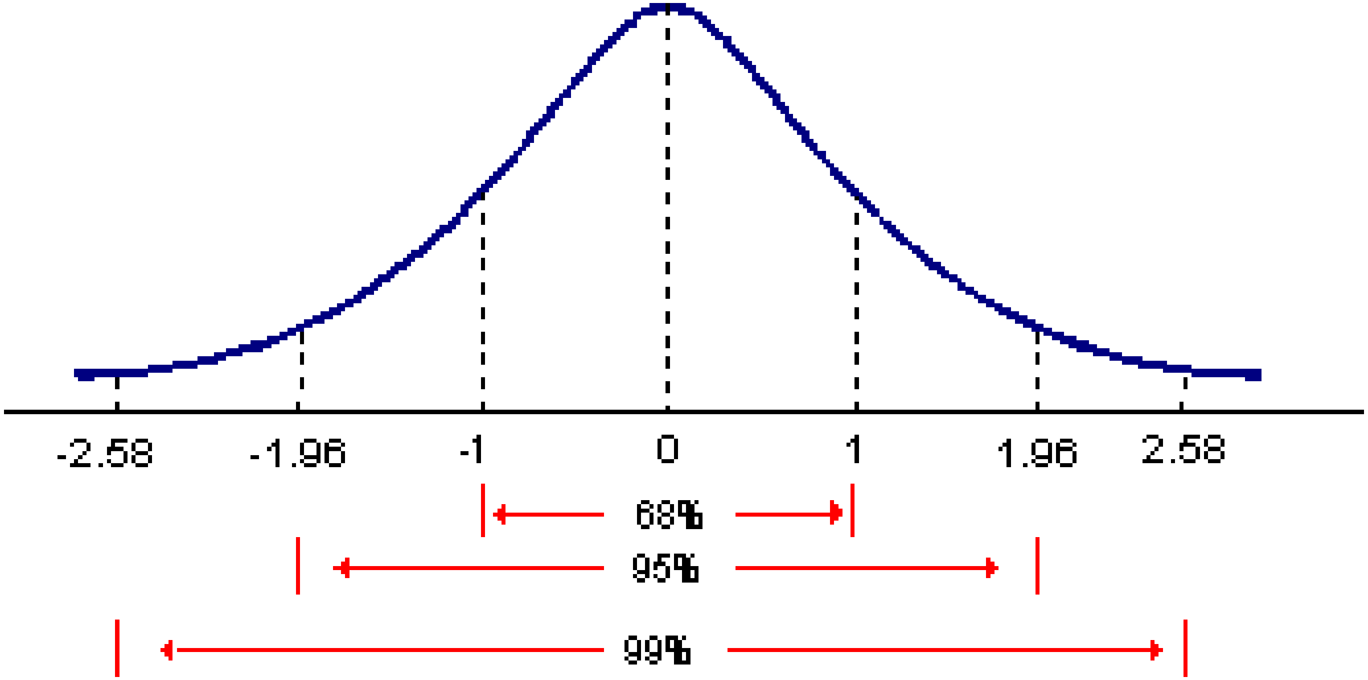
X=70

- $Z=(70-50)/5=20/5=4$



X=70 cntd.

- *4>1.96;* falls outside 95% confidence interval
- *4>2.58;* falls outside 99% confidence interval



What is the probability of the observation?

- Actual value, $P(X=70) = 0.5^{70} \times 0.5^{30}$
- By Z-score method:
- $P(X=70)=P(69.5<=X<=70.5)$
- $\rightarrow P[(69.5-50)/5<=Z<=(70.5-50)/5]$
- $=P(3.9<=Z<=4.1)$
- $=Z_{4.1}-Z_{3.9}$
- *Very small!*
- *P(X)* is called the *p-value*

Read the Z-score table

Standard Normal Probabilities

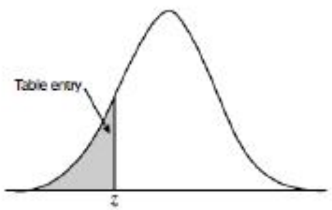


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard Normal Probabilities

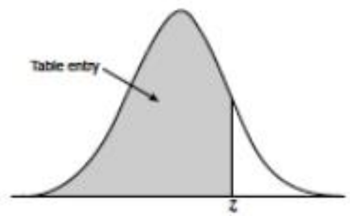
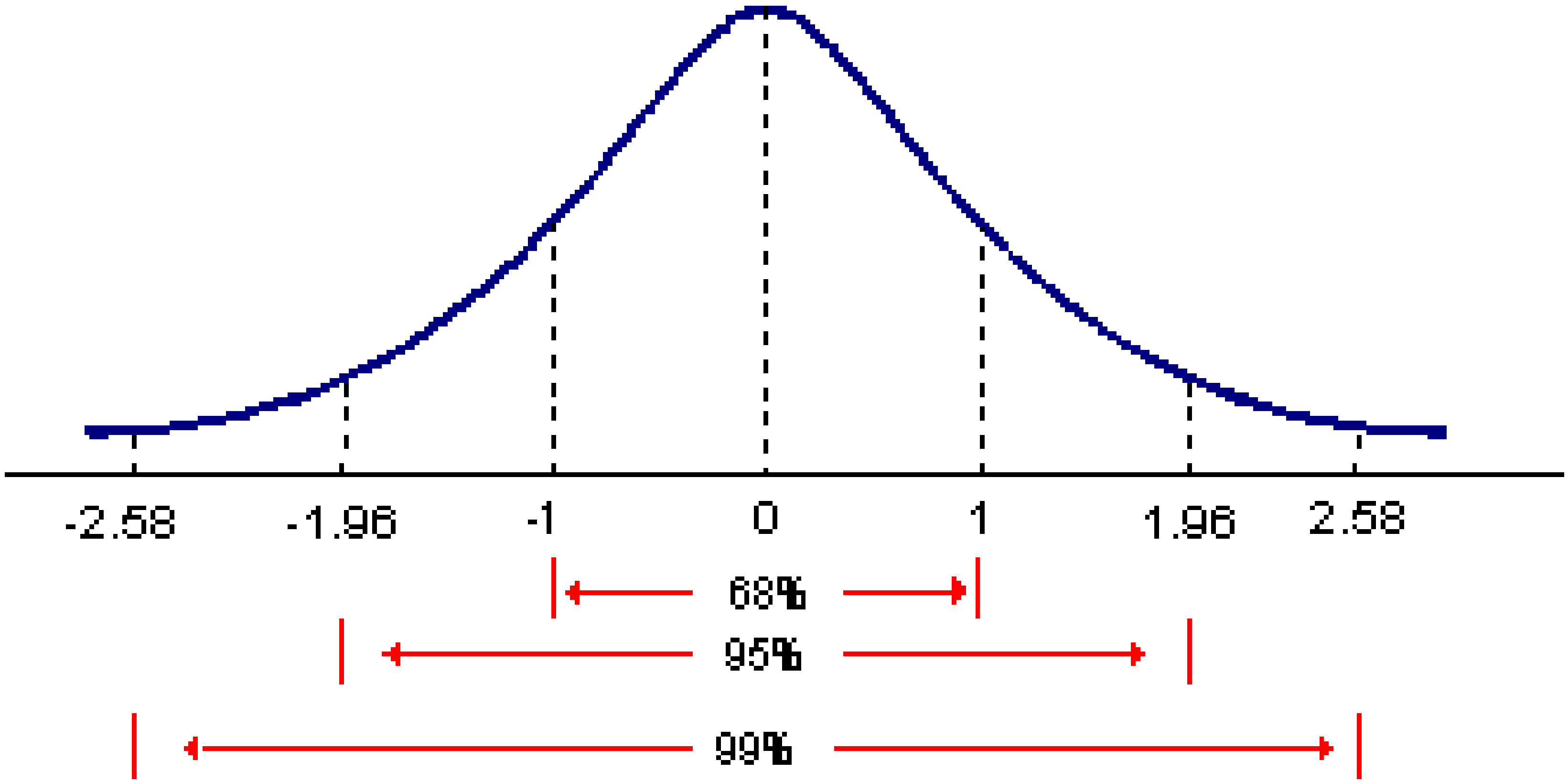


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

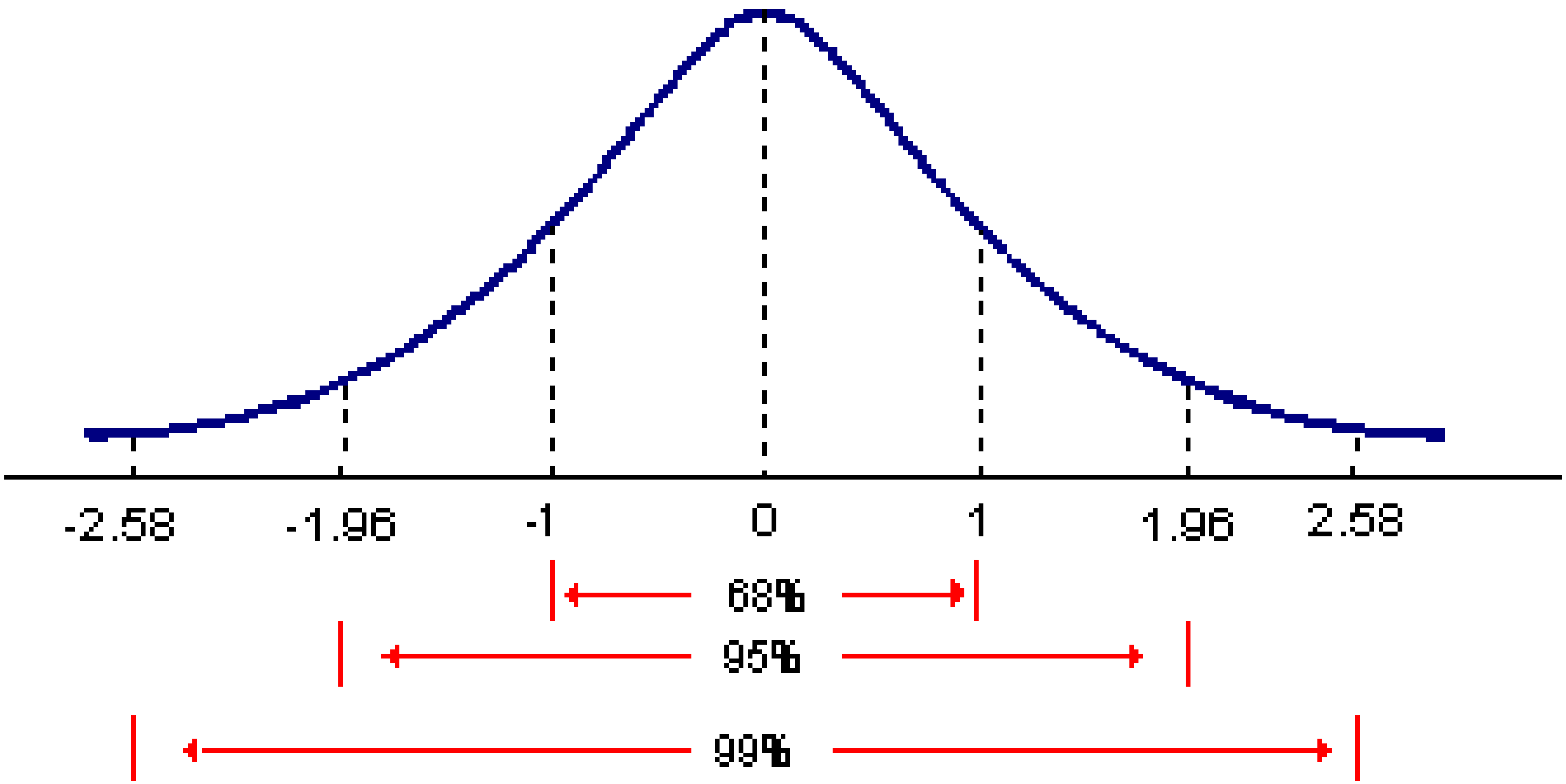
X=60

- *Z=(60-50)/5=10/5=2; can reject for 95% CI, but not for 99%*



X=55

- *Z=(55-50)/5=5/5=1; cannot reject under 95% CI*



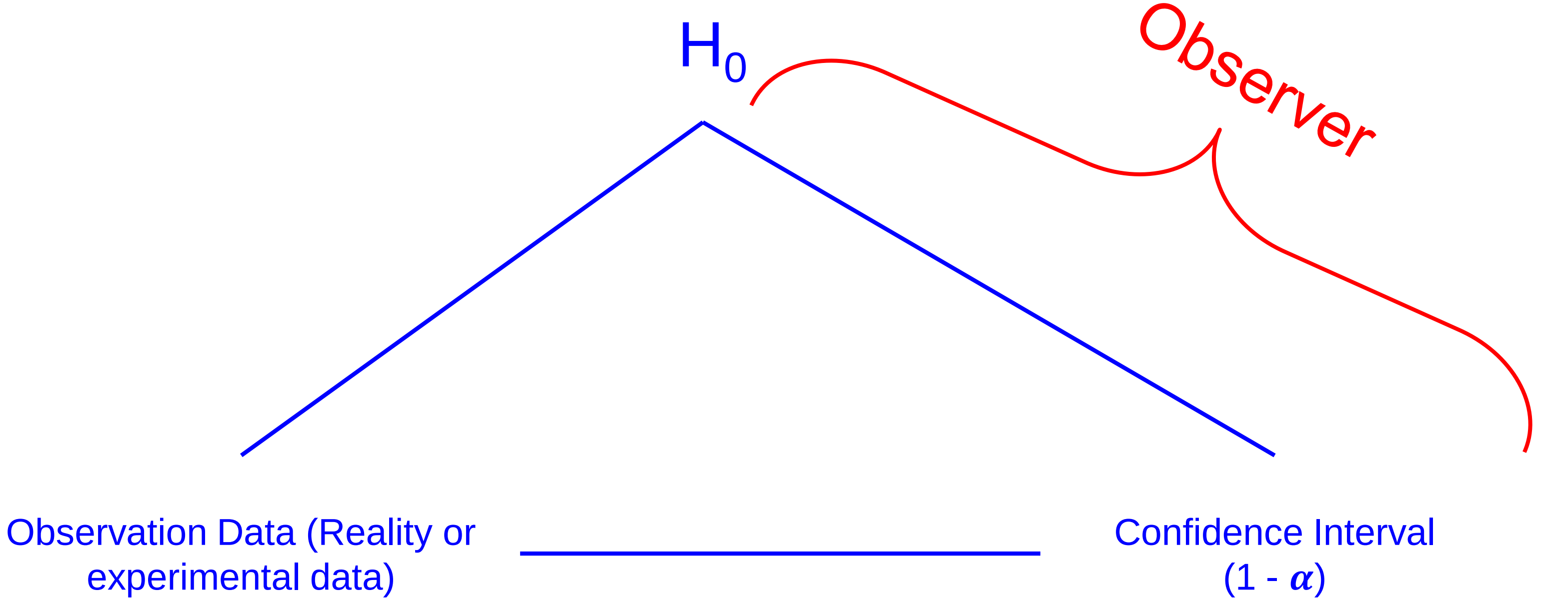
P-values for X=60

- Actual value, $P(X=60)= 0.5^{60} X 0.5^{40}$
- $P(X=55)= 0.5^{55} X 0.5^{45}$
- By Z-score method:
- *$P(X=60)=P(59.5<=X<=61.5)$*
- ➔*$P[(59.5-50)/5<=Z<=(60.5-50)/5]$*
- *$=P(1.9<=Z<=2.1)$*
- *$=Z_{2.1}-Z_{1.9}$*
- *$=0.9821-0.9713=0.0108=0.01<0.05$*

P-values for X=55

- Actual value, $P(X=55)= 0.5^{55} \text{ X } 0.5^{45}$
- By Z-score method:
- $P(X=55)=P(54.5<=X<=55.5)$
- $\rightarrow P[(54.5-50)/5<=Z<=(55.5-50)/5]$
- $=P(0.9<=Z<=1.1)$
- $=Z_{1.1}-Z_{0.9}$
- $=0.8643-0.8159=0.04$ *approx*

Hypothesis testing through triangulation



Digression: Hypothesis Testing in Logic

Using Predicate Calculus

Himalayan Club example

- Introduction through an example *(Zohar Manna, 1974):*
  - Problem: A, B and C belong to the Himalayan club. Every member in the club is either a mountain climber or a skier or both. A likes whatever B dislikes and dislikes whatever B likes. A likes rain and snow. No mountain climber likes rain. Every skier likes snow. *Is there a member who is a mountain climber and not a skier?*
- Given knowledge has:
  - Facts
  - Rules

Example contd.

- Let *mc* denote mountain climber and *sk* denotes skier. Knowledge representation in the given problem is as follows:
  1. *member(A)*
  2. *member(B)*
  3. *member(C)*
  4. $\forall x[member(x) \rightarrow (mc(x) \vee sk(x))]$
  5. $\forall x[mc(x) \rightarrow \sim like(x,rain)]$
  6. $\forall x[sk(x) \rightarrow like(x, snow)]$
  7. $\forall x[like(B, x) \rightarrow \sim like(A, x)]$
  8. $\forall x[\sim like(B, x) \rightarrow like(A, x)]$
  9. *like(A, rain)*
  10. *like(A, snow)*
  11. *Question:* $\exists x[member(x) \wedge mc(x) \wedge \sim sk(x)]$
- We have to infer the 11th expression from the given 10.
- Done through Resolution Refutation.

Club example: Inferencing

1. $member(A)$
2. $member(B)$
3. $member(C)$
4. $\forall x[member(x) \rightarrow (mc(x) \vee sk(x))]$
   - Can be written as
   $[member(x) \rightarrow (mc(x) \vee sk(x))]$
   - $\sim member(x) \vee mc(x) \vee sk(x)$
5. $\forall x[sk(x) \rightarrow lk(x, snow)]$
   - $\sim sk(x) \vee lk(x, snow)$
6. $\forall x[mc(x) \rightarrow \sim lk(x, rain)]$
   - $\sim mc(x) \vee \sim lk(x, rain)$
7. $\forall x[like(A, x) \rightarrow \sim lk(B, x)]$
   - $\sim like(A, x) \vee \sim lk(B, x)$

8. $\forall x[\sim lk(A,x) \rightarrow lk(B,x)]$

– $lk(A,x) \vee lk(B,x)$

9. $lk(A, rain)$

10. $lk(A, snow)$

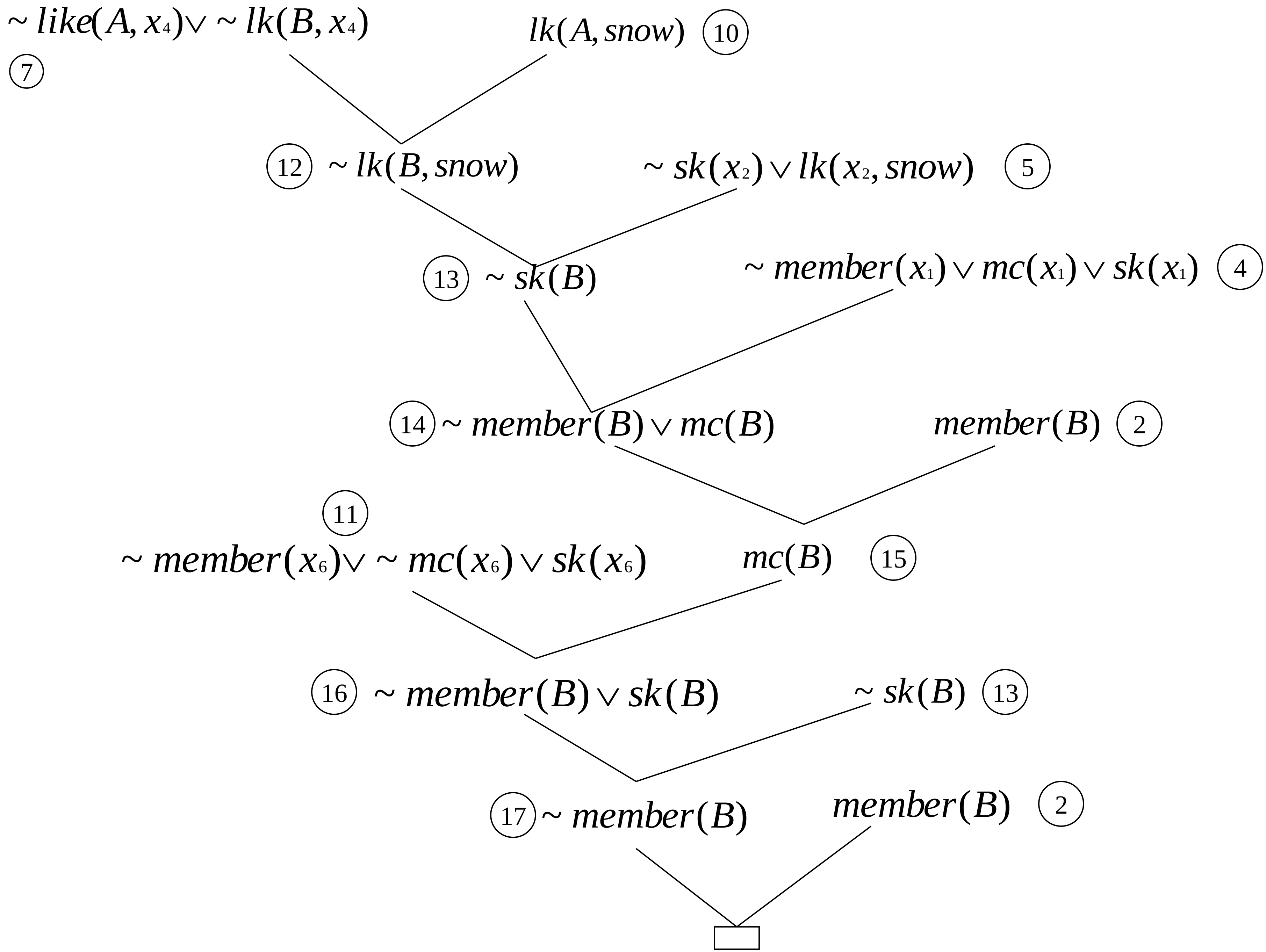
11. $\exists x[member(x) \wedge mc(x) \wedge \sim sk(x)]$

$\forall x[\sim member(x) \vee \sim mc(x) \vee sk(x)]$

– Negate–

- Now standardize the variables apart which results in the following

1. $member(A)$
2. $member(B)$
3. $member(C)$
4. $\sim member(x_1) \vee mc(x_1) \vee sk(x_1)$
5. $\sim sk(x_2) \vee lk(x_2, snow)$
6. $\sim mc(x_3) \vee \sim lk(x_3, rain)$
7. $\sim like(A, x_4) \vee \sim lk(B, x_4)$
8. $lk(A, x_5) \vee lk(B, x_5)$
9. $lk(A, rain)$
10. $lk(A, snow)$
11. $\sim member(x_6) \vee \sim mc(x_6) \vee sk(x_6)$


~ like(A, x4) ∨ ~ lk(B, x4) 7
lk(A, snow) 10
12 ~ lk(B, snow)
~ sk(x2) ∨ lk(x2, snow) 5
13 ~ sk(B)
~ member(x1) ∨ mc(x1) ∨ sk(x1) 4
14 ~ member(B) ∨ mc(B)
member(B) 2
11
~ member(x6) ∨ ~ mc(x6) ∨ sk(x6)
mc(B) 15
16 ~ member(B) ∨ sk(B)
~ sk(B) 13
17 ~ member(B)
member(B) 2

Null Hypothesis: H_0

- H_0: The club does NOT have any member that is a mountain climber (MC) and not a skier (SK)

- Key question: Under H_0, is the observation valid?

- In other words: is the hypothesis consistent with the data?

Methodology

- If Hypothesis not consistent with data, hypothesis must be rejected

- Data cannot be rejected

- Data is GOLD!

Data for Himalayan Club Example

(1) A, B and C belong to the Himalayan club.
(2) Every member in the club is either a mountain climber or a skier or both.
(3) A likes whatever B dislikes and
(4) dislikes whatever B likes.
(5) A likes rain and snow.
(6) No mountain climber likes rain.
(7) Every skier likes snow

Null Hypothesis for Himalayan Club Example

- H_0: *There is NOT a single member who is a mountain climber and not a skier*
- H_0 inconsistent with data
- So must be rejected
- Methodology: Logical Inferencing-Resolution-Refutation

Interval Estimate

Sample Mean and Population Mean

- $X_1, X_2, X_3, \ldots, X_n$ is a sample from a normal distribution having unknown mean μ and known variance σ^2.
- Maximum likelihood point estimator of μ is

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$\bar{X}$

- We know that $\bar{X}$ is normally distributed with mean μ and known standard deviation $\sigma/\sqrt{n}$
- So the following is standard normal distribution:

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

95% confidence interval

$$P\left[-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96\right] = 0.95$$

$$\Rightarrow P\left[-1.96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

$$\Rightarrow P\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

A Manufacturing situation

Suppose that a machine part manufacturer has made parts with the dimensions as given below:

(a) 9 pieces of the machine part

(b) dimensions are respectively

5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5

Suppose somehow it is known that IF the parts could be measured on the whole population sample by sample, the variance of the measurement of dimensions would be 4

(*artificial? Yes, but useful for concept building*)

95% confidence interval for μ

$5+8.5+12+15+7+9+7.5+6.5+10.5=81$

$\bar{X}=81/9=9$

It follows that under the assumption that the values are independent, a 95% confidence interval for μ is

$[9-1.96.(2/3), 9+1.96.(2/3)]$

$=$**(7.6, 10.31)**

Interpretation of the observation

Based on the

(a) observation (9 samples)

(b) knowledge obtained somehow that variance is 4

We reach the 95% confidence interval as (7.69, 10.31)

Qualitatively

- If the manufacturer says, “I can ensure a part dim of 15”, we cannot trust!!

- If the maker says, “I assure dim of 10”, we CAN trust

- Trust with 95% confidence

Two sided and one sided confidence intervals

- What we saw is 2 sided confidence interval

- Similarly, one sided upper and lower confidence intervals

95% one sided intervals

- Upper

$$\left(\bar{X} - 1.645 \frac{\sigma}{\sqrt{n}}, \infty \right)$$

- Lower

$$\left(-\infty, \bar{X} + 1.645 \frac{\sigma}{\sqrt{n}} \right)$$

Type I and Type II error

- Type I: incorrectly reject H_0, when it should have been accepted. H_0 is actually true in the population

- Type II: incorrectly accept H_0 when it should have been rejected. H_0 is actually false in the population.

Type-I and Type-II errors: **Always wrt Null Hypothesis H_0**

actual \ as per data	ACCEPT	REJECT
TRUE	***No Error***	***Type- I error (false –ve)***
FALSE	***Type-II Error (false +ve)***	***No Error***

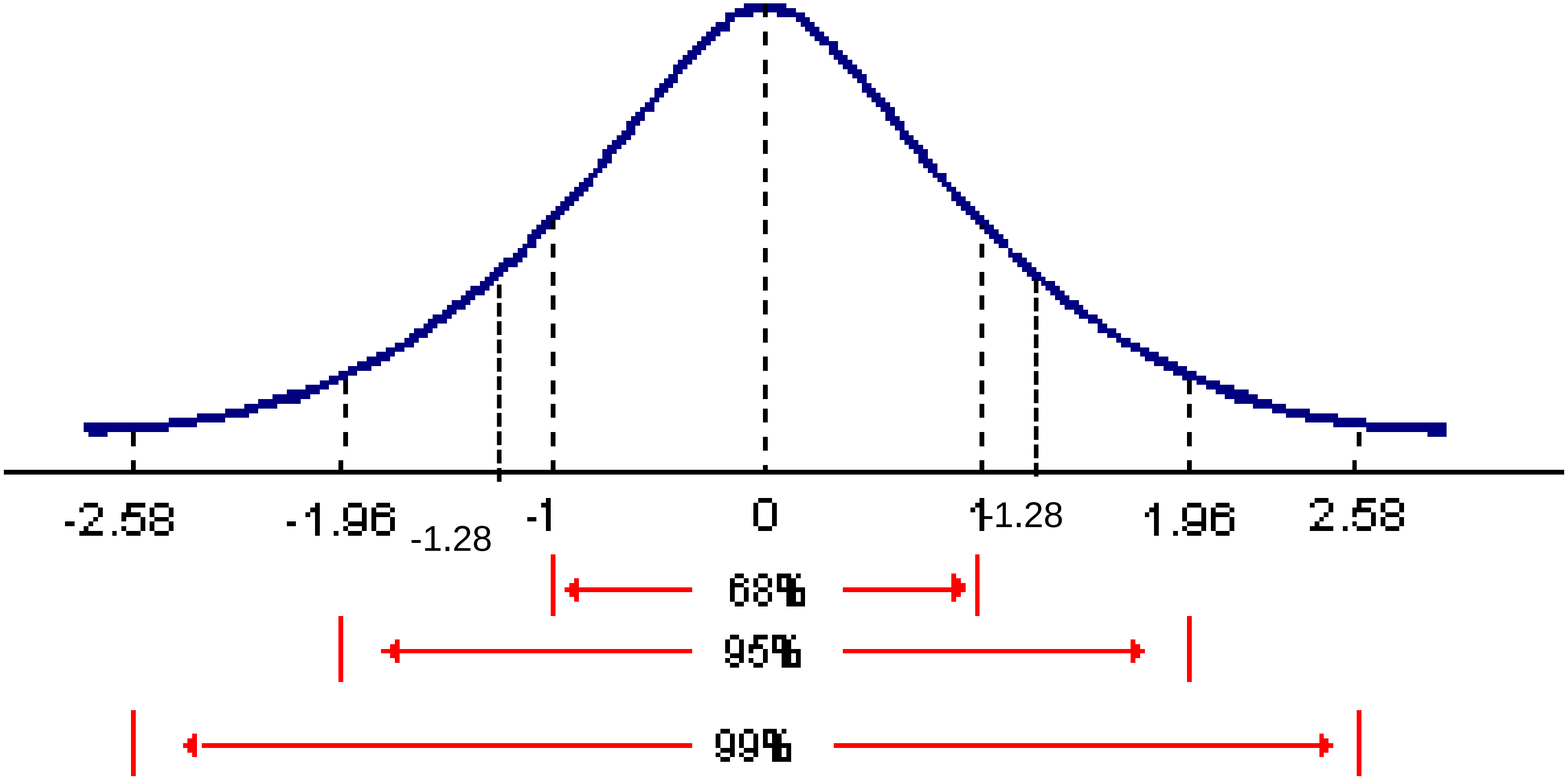
More on H_0

- **Data**: (a) All men are mortal, (b) Shakespeare is a man
- **H_0**: Shakespeare is not mortal
- **Contradiction**: Shakespeare is not a man
- **Conclusion**: Reject H_0

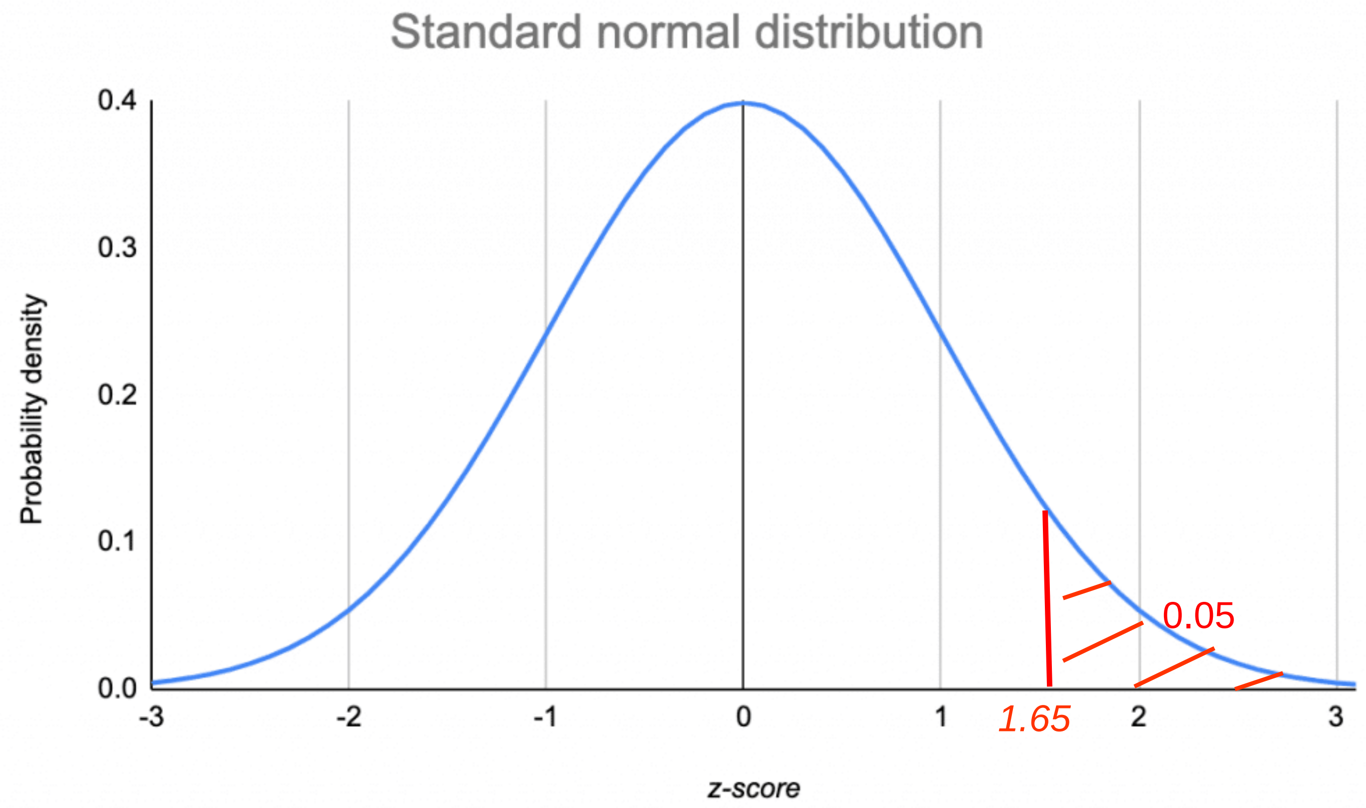
A useful table

test-type (col)			
vs.	Two-Tail	1 sided to +inf	1 sided from -inf
Confidence Interval (significance level)			
90% (0.10)	(- and +) 1.65	-1.28 to +inf	-inf to +1.28
95% (0.05)	(- and +) 1.96	-1.65 to +inf	-inf to +1.65
99% (0.01)	(- and +) 2.58	-2.33 to +inf	-inf to 2.33

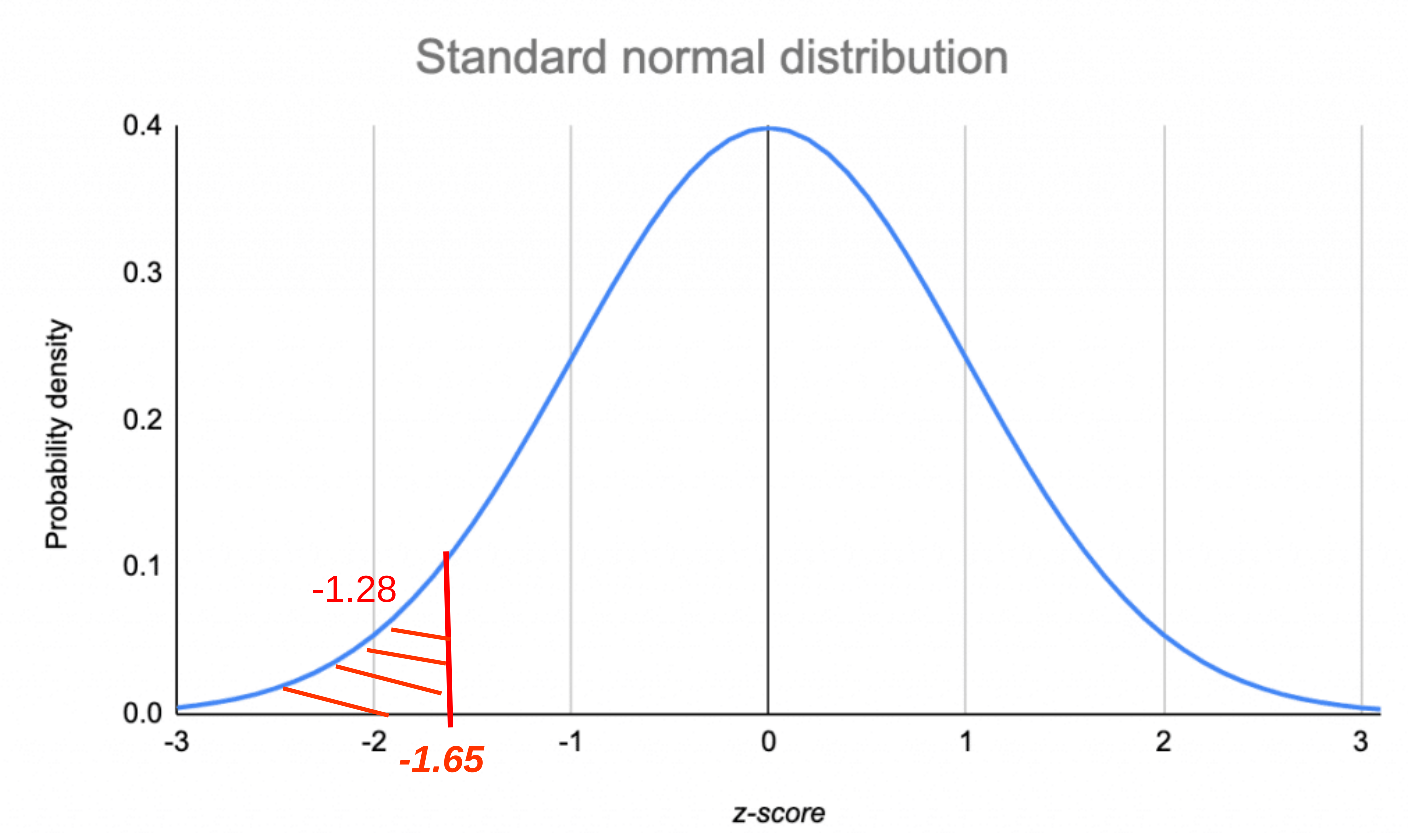
2 sided 95% confidence interval



95% 1-sided confidence interval (from –inf)



95% 1-sided confidence interval (to +inf)



Illustration

Problem Statement: bottling of fluid

- A factory has a machine that- the factory claims- dispenses 80mL of fluid in a bottle. This needs to be tested. A sample of 40 bottles is taken. The average amount of fluid is 78mL with standard deviation of 2.5. Verify the factory’s claim.

https://www.youtube.com/watch?v=zJ8e_wAWUzE

Essential elements (1/n)

- Examine the problem carefully, read the problem statement again and again, discuss the issues threadbare
- ***1. Formulate H_0***
- 2. Formulate H_A
- 3. *Decide confidence interval* (usually 90, 95 or 99%)→ this automatically fixes the significance level (0.10, 0.05 or 0.01)

This sets the rule of the game, cannot change going forward !!

Essential elements (2/n)

- From H_0 and H_A, decide 2-sided or 1 sided test
  - 2 sided for '=' or '≠'
  - 1 sided for '>=' or '=<'
    - Upper (right) for '>='
    - Lower (left) for '=<'

Essential elements (3/n)

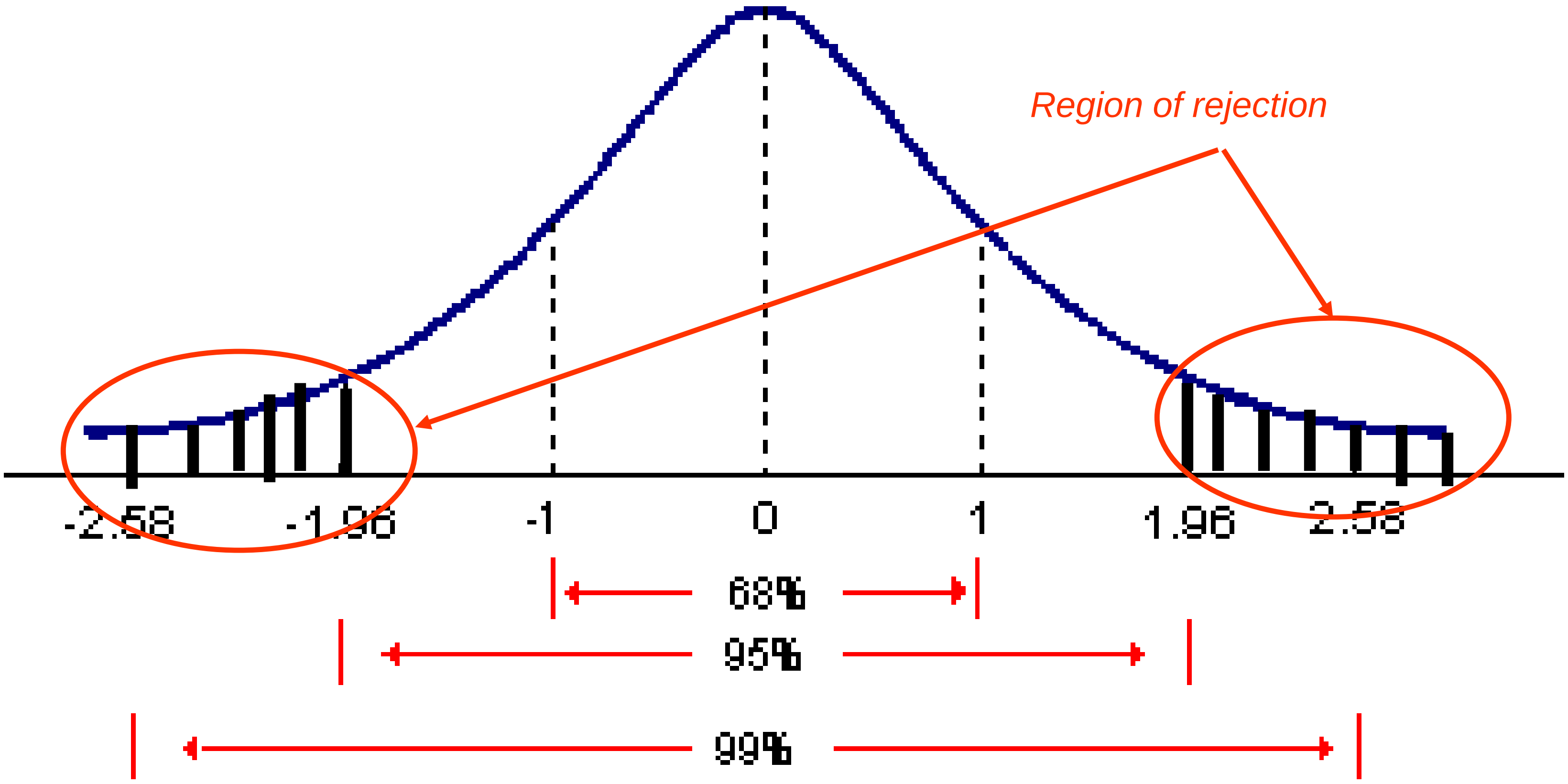
- Decide Z-test/T-test/F-test/ChiSquare test
- If Z-test, Z_c (critical value)=

  +- 1.65 for 90% confidence interval, + 1.28 for upper 90% confidence interval, - 1.28 for lower 90% confidence interval

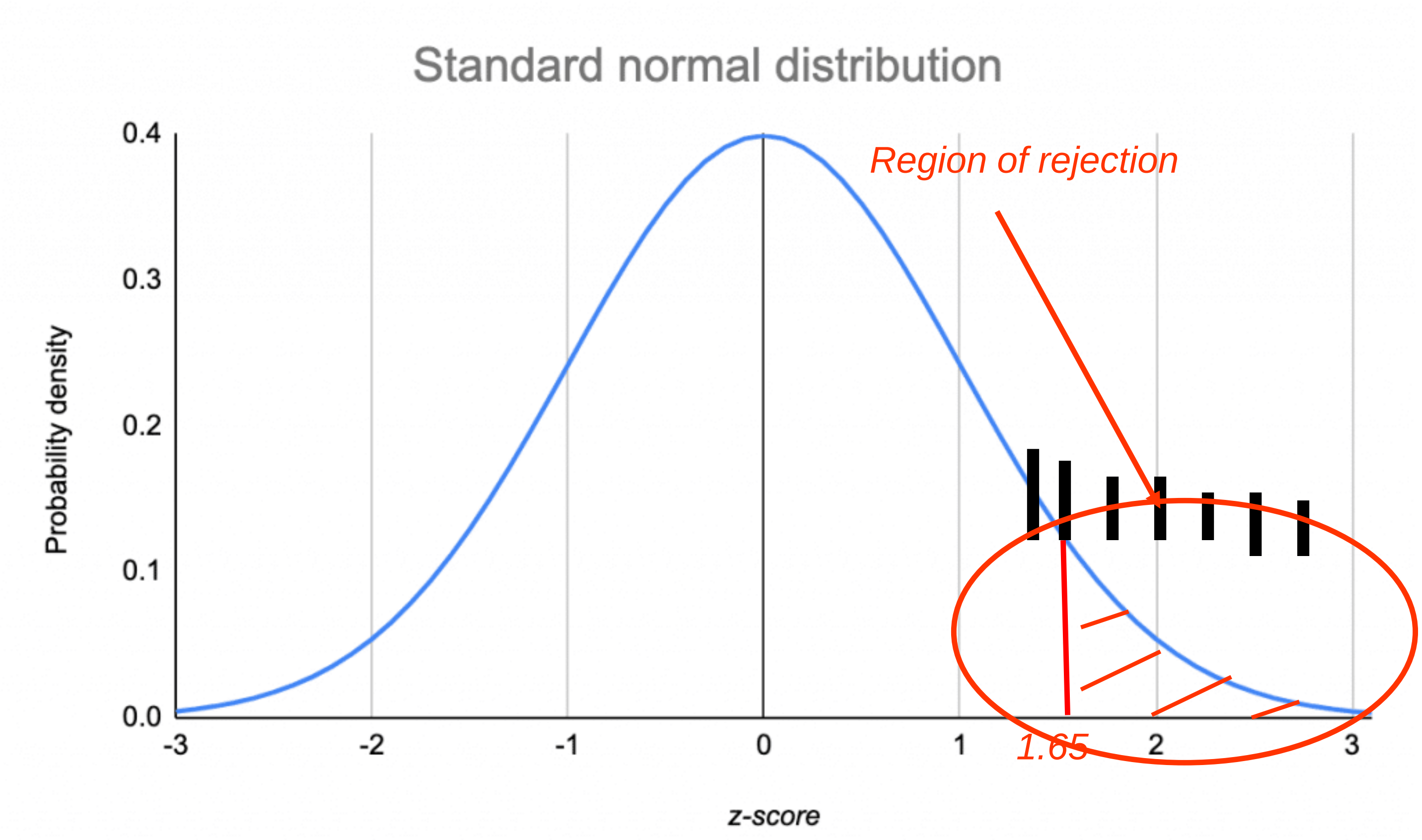
  +- 1.96 for 95% confidence interval, +1.65 for upper 95% confidence interval, -1.65 for lower 95% confidence interval

  +-2.58 for 99% confidence interval, +2.33 for upper

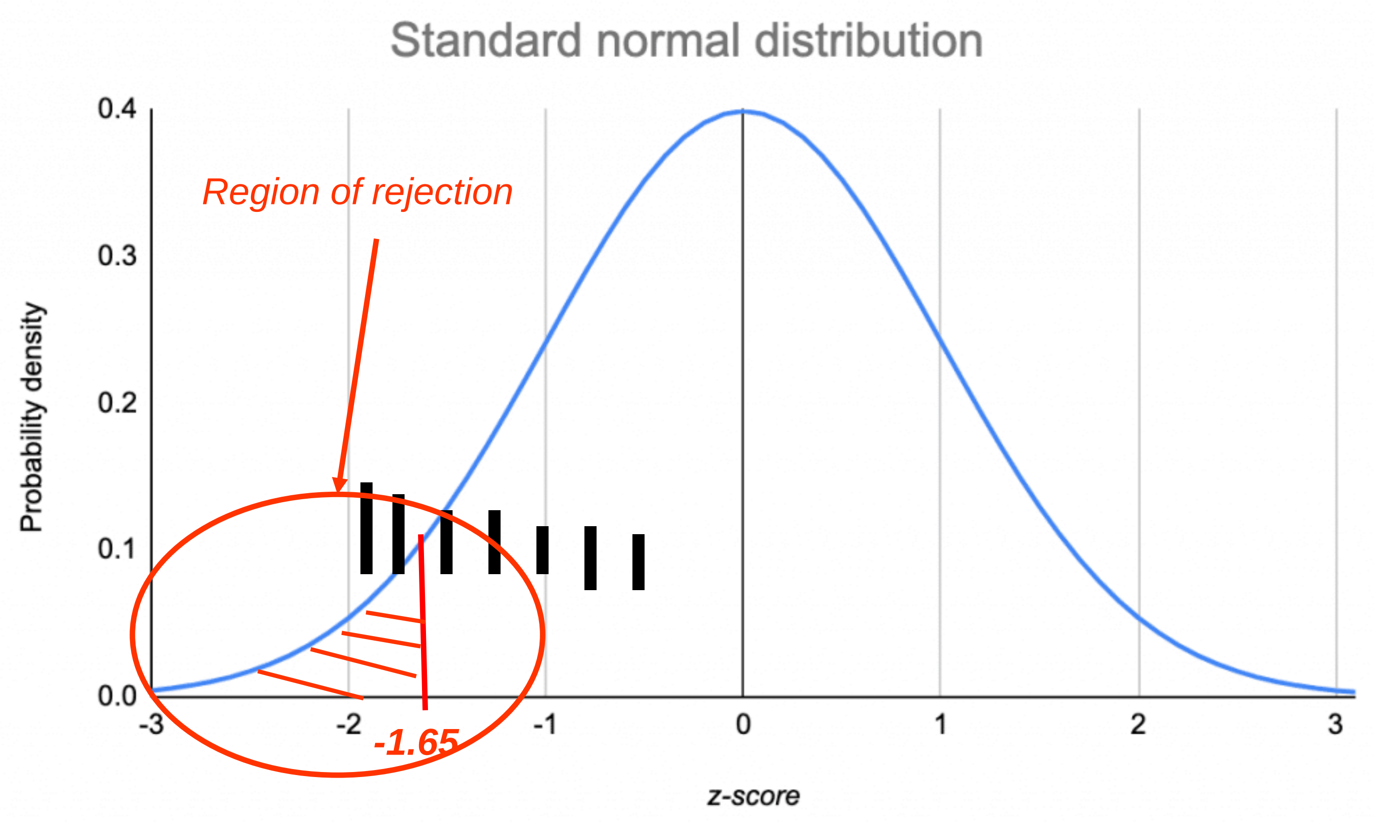
2 sided 95% confidence interval



1-sided confidence interval (upper/right)



1-sided confidence interval (lower/left)



Essential elements (4/n)

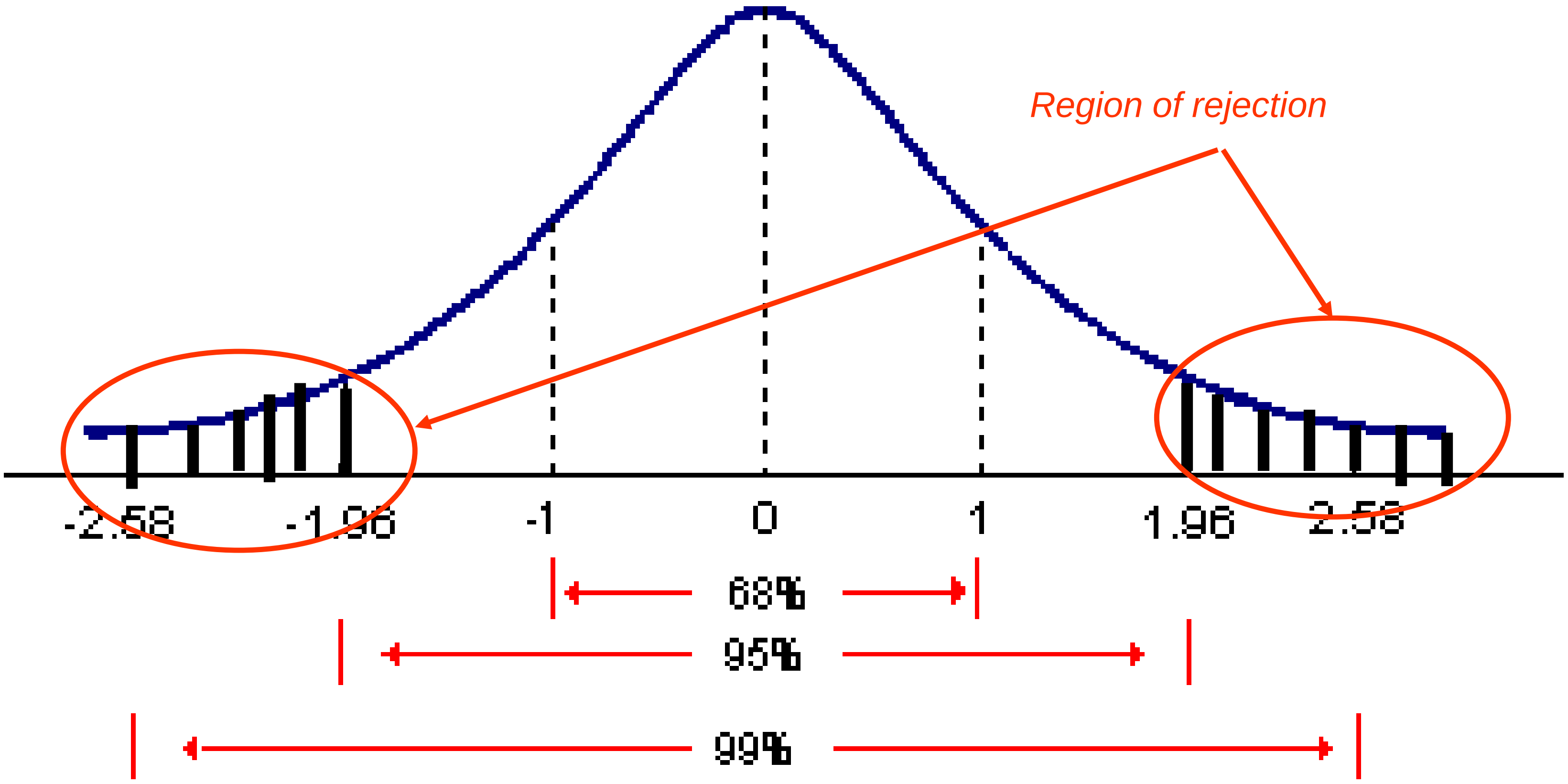
- **Under H0**, determine test statistic from the data called OBSERVED

- For Z-test, Z_o (observed)
- If Zo is inside the "REJECT" region, reject H_0
- Else cannot reject H_0

Back to Illustration: bottling of fluid

- Claimed population mean, μ=80
- n=40, sample mean, μ_{obs}=78, sample standard deviation, σ_{obs}=2.5
- H_0: μ=80
- H_A: 3 options
  - $\mu \neq 80$ (2-tailed or 2-sided test)
  - $\mu >= 80$ (1-tailed Upper)
  - $\mu <= 80$ (1-tailed Lower)

2 sided 95% confidence interval



2-tailed analysis

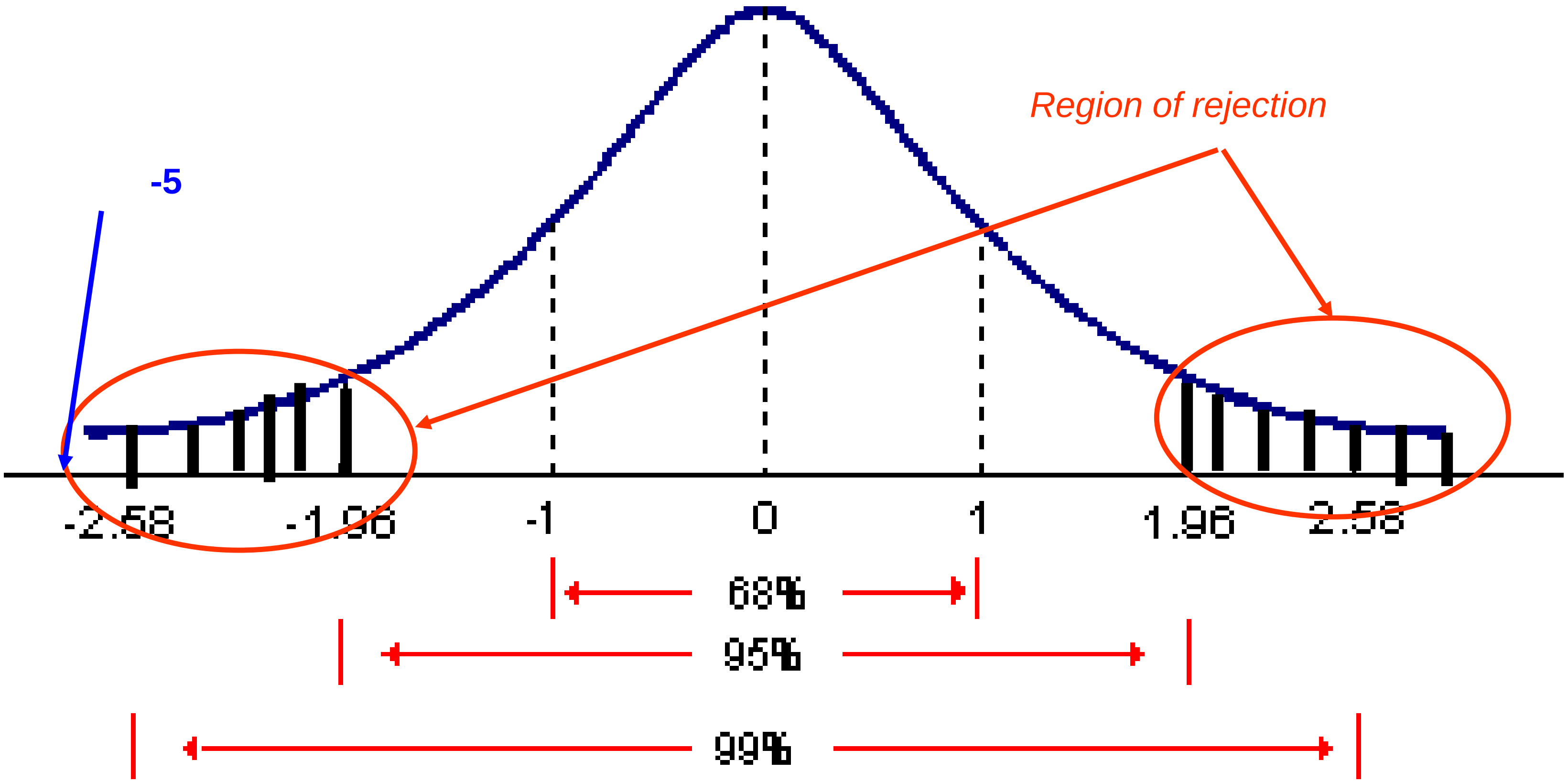
- Z_c = +- 1.96
- Z_{obs} =

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{78 - 80}{\frac{2.5}{\sqrt{40}}}$$

$$= -5\,(approx.)$$

- Falls in rejection region

2 sided 95% confidence interval



Z-test based observation (2-tailed)

- $-5 < -1.96$
- We reject the null hypothesis
- The claim that the machine fills bottles with 80mL fluid is rejected based on the evidence

99% confidence interval

- -5 still in rejection region
- $-5.0 < -2.58$
- So for 99% confidence interval also the hypothesis is rejected

90% confidence interval

- -5 still in rejection region
- $-5.0 < -1.28$
- So for 90% confidence interval also the hypothesis is rejected

Another problem on HT

Problem statement

An NLP tool is built, which given a corpus of sentences, identifies the nouns and tags them as 1 and the rest of the words as 0. For example, given the sentence “Mumbai is a big city”, the tags produced are “Mumbai_1 is_0 a_0 big_0 city_1”. It is required to assess if the tool is doing anything better than random guessing.

Binomial Distribution

Given that *p* is the probability of success, i.e., getting the correct tag, the probability of getting *K* correct tags in a corpus sample of *N* words can be modelled as a binominal distribution and its value is equal to $\boldsymbol{{}_{N}C_{K}\,p^{K}(1-p)^{N-K}}$, where ${}_{N}C_{K}$ indicates *N* choose *K*

Mean and S.D. of Binomial Distribution

The mean and the standard deviation of the binomial distribution are respectively, ***Np*** and ***sqrt[Np(1-p)]***

Normal approx. to Binomial

Now, it is known that binomial distribution can be approximated by normal distribution. The conditions for this are ***Np>5*** and ***N(1-p)>5***

Statement with 95% confidence

The NLP tool is given a corpus of 100 words (consider punctuations also as words). The range of values of *K* (from part A) should lie in the range from ____ to ______for us to be able to state that “the tool is doing random guessing with no preference for correct vs. incorrect tag, and I am 95% confident in my statement”.

- **Ans**: **from 41 to 59, both inclusive**

Justification

mean=100 X 0.5=50,

standard deviation=sqrt(100 X 0.5 X 0.5)=5

Z=(X-50)/5

Putting Z=-1.96 and +1.96 and taking the ceiling of lower value and floor of the upper value, we get the answer.