Statistical Data Analysis

Prof. Dr. Nizamettin AYDIN

naydin@yildiz.edu.tr

http://www.yildiz.edu.tr/~naydin

Examples

Example 1

- We have measured the height (in inches) and weight (in pounds) for five newborn babies (shown in the table).

Observation	Height	Weight
1	18	7.8
2	21	9.1
3	17	8.2
4	16	6.4
5	19	8.8

- Manually calculate the mean and standard deviation of height and weight; show all the steps.
- Table: Height (in inches) and weight (in pounds) for five newborn babies

Answer 1

$$Mean\ of\ Height = \bar{x}_H = \frac{x_{H1} + x_{H2} + \cdots + x_{Hn}}{\#\ of\ observations}$$

$$= \frac{18 + 21 + 17 + 16 + 19}{5} = 18.2$$

$$Mean\ of\ Weight = \bar{x}_W = \frac{x_{W1} + x_{W2} + \cdots + x_{Wn}}{\#\ of\ observations}$$

$$= \frac{7.8 + 9.1 + 8.2 + 6.4 + 8.8}{5} = 8.06$$

Answer 1

$$SD\ of\ Height = \sqrt{\sum_{i=1}^{n} \frac{(x_{Hi} - \bar{x}_H)^2}{n-1}} =$$

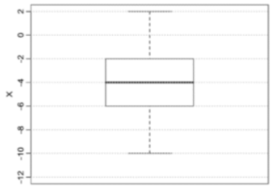
$$\sqrt{\frac{(18 - 18.2)^2 + \cdots + (19 - 18.2)^2}{5 - 1}} = 1.92$$

$$SD\ of\ Weight = \sqrt{\sum_{i=1}^{n} \frac{(x_{Wi} - \bar{x}_W)^2}{n-1}}$$

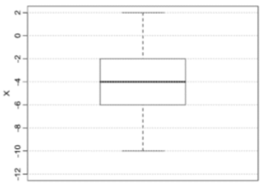
$$= \sqrt{\frac{(7.8 - 8.06)^2 + \cdots + (8.8 - 8.06)^2}{5 - 1}} = 1.06$$

Example 2

- Based on the following boxplot, write down the five-number data summary, range and IQR of variable X.
- Boxplot of variable X



Answer 2



Five number summary =

Min = -10,
Q1 = -6,
Q2 (median) = -4,
Q3 = -2,
Max = 2

Range = Max – Min
= 2 - (-10) = 12

IQR = Q3 – Q1
= -2 - (-6) = 4

Example 3

- Some values of a row in an image matrix is given as;
  x = {..., 4, 6, 50, 7, 5, 80, 4, 4, 7, 60, 6, 4, 8, 40, 7, 5, ...} .
- Inspecting the values in this row reveals that this image is contaminated by **impulsive noise** (for example 50, 80, …).

a. Suggest a filtering method to remove this impulsive noise.

b. What will be the resulting values of the row after applying the filter?

Answer 3

a. I would suggest to use **moving (sliding) median filter**,
as it is a non-linear filter to remove impulsive noise.

b.

x_{new} = {..., 4, 6, 7, 7, 7, 5, 4, 4, 7, 7, 6, 6, 8, 8, 7, 5, ...}

x_{old} = {..., 4, 6, 50, 7, 5, 80, 4, 4, 7, 60, 6, 4, 8, 40, 7, 5, ...}

Example 4

- A large drug company has 100 potential new prescription drugs under clinical test.
- About 20% of all drugs that reach this stage are eventually licensed for sale.
- What is the probability that at least 15 of the 100 drugs are eventually licensed?
  - Assume that the binomial assumptions are satisfied, and use a normal approximation with continuity correction.

Answer 4

- The mean of y
  $\mu = n\theta = 100\times0.2 = 20$
- The standard deviation
  $\sigma = \text{sqrt}(n\theta(1-\theta)) = \text{sqrt}(100\times0.2\times(1-0.2)) = 4$
- The desired probability is that 15 or more drugs are approved.
- Because y =15 is included, the continuity correction is to take the event as y greater than or equal to 14.5.

$$P(y \geq 14.5) = P\left(z \geq \frac{14.5-20}{4.0}\right) = P(z \geq -1.38) = 1 - P(z < -1.38)$$
$$= 1 - .0838 = .9162$$

Continuity Correction Factor

- used when a continuous probability distribution is used to approximate a discrete probability distribution.
  - For example, when you want to use the normal to approximate a binomial.
- According to the Central Limit Theorem, the sample mean of a distribution becomes approximately normal if the sample size is “large enough.”
  - For example, the binomial distribution can be approximated with a normal distribution as long as n×p and n×q are both at least 5. Here,
    - n = how many items are in your sample,
    - p = probability of an event (e.g. 60%),
    - q = probability the event doesn’t happen (100% – p).

Continuity Correction Factor

- The continuity correction factor accounts for the fact that a normal distribution is continuous, and a binomial is not.
- When you use a normal distribution to approximate a binomial distribution, you're going to have to use a continuity correction factor.
- It is as simple as adding or subtracting 0.5 to/from the discrete x-value:
  - use the following table to decide whether to add or subtract.
    - If P(X=n) use P(n – 0.5 < X < n + 0.5)
    - If P(X > n) use P(X > n + 0.5)
    - If P(X ≤ n) use P(X < n + 0.5)
    - If P (X < n) use P(X < n – 0.5)
    - If P(X ≥ n) use P(X > n – 0.5)

13

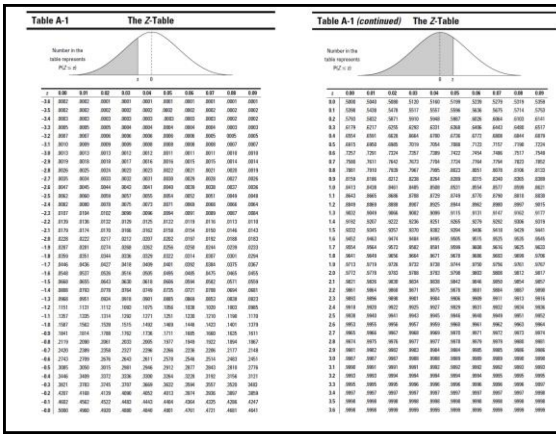


Table A-1 The Z-Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
–3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0003	.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

Table A-1 (continued) The Z-Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

14

Finding probabilities for Z with the Z-table

- To use the *Z*-table to find probabilities for the standard normal (*Z*-) distribution
  - Go to the row that represents the first digit of your z-value and the first digit after the decimal point.
  - Go to the column that represents the second digit after the decimal point of your z-value.
  - Intersect the row and column.
- This result represents p(Z < z), the probability that the random variable Z is less than the number z (also known as the percentage of z-values that are less than yours).

15

Finding Probabilities for a Normal Distribution

- Draw a picture of the distribution.
- Translate the problem into one of the following:
  - $p(X < a)$, $p(X > b)$, or $p(a < X < b)$.
  - Shade in the area on your picture.
- Standardize *a* (and/or *b*) to a z-score using the z-formula:

$$z = \frac{x-\mu}{\sigma}$$

- Look up the z-score on the Z-table and find its corresponding probability.
  - If you need a "less-than" probability — that is, $p(X < a)$ — you're done.
  - If you want a "greater-than" probability — that is, $p(X > b)$ — find $1 - p(X < b)$.
- If you need a "between-two-values" probability — that is, $p(a < X < b)$ — perform the same steps defined above for *b* and *a*, and subtract the results.

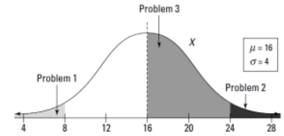
16

Example 5

- Suppose that you enter a fishing contest.
- The contest takes place in a pond where the fish lengths have a normal distribution with mean μ = 16 cm and standard deviation σ = 4 cm.
  1. What is the chance of catching fish less than 8 cm?
  2. Suppose a prize is offered for any fish over 24 cm. What is the chance of winning a prize?
  3. What is the chance of catching a fish between 16 and 24 cm?

17

Answer 5



- Need to find
  - $p(X < 8)$,
  - $p(X > 24)$,
  - $p(16 < X < 24)$.
- First change the x-values to z-values using the z-formula: $z = \frac{x-\mu}{\sigma}$
  1. chance of catching fish less than 8 cm:

$$p(X<8)=p\left(Z<\frac{8-16}{4}\right)=p(Z<-2) = 0.0228$$

  2. chance of winning a prize

$$p(X>24)=p\left(Z>\frac{24-16}{4}\right)=p(Z>2)$$

18

Answer 5

$p(Z > 2.00) = 1 – p(Z < 2.00) = 1 – 0.9772 = 0.0228$

3. chance of catching a fish between 16 and 24 cm

$$p(16 < X < 24) = p\left(\frac{16-16}{4} < Z < \frac{24-16}{4}\right) = p(0 < Z < 2)$$

- find $p(Z < 2.00)$, which is 0.9772
- find $p(Z < 0)$, which is 0.5000

$p(0 < Z < 2) = 0.9772 - 0.5000 = 0.4772$

- The chance of a fish being between 16 and 24 cm is 0.4772

Example 6

- We assume that the probability distribution of blood pressure, X, is $N(\mu,\sigma^2)$distribution.
  - Suppose we know that $\sigma = 6$.
  - To estimate μ, we randomly selected 9 people and measured their blood pressure.
  - The sample mean is $\bar{x} = 110$.

a. Write down the sampling distribution of the sample mean $\overline{X}$ and find its standard deviation.

b. Find the 80% confidence interval estimation for μ.

Answer 6

a. $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$, $\quad \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{9}} = \frac{6}{3} = 2$

b. With the point estimate $\bar{x}$, the confidence interval for the population mean at c confidence level is

$$\left[\bar{x} - z_{crit} \times \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{crit} \times \frac{\sigma}{\sqrt{n}}\right]$$
$$= \left[110 - z_{crit} \times \frac{\sigma}{\sqrt{n}}, 110 + z_{crit} \times \frac{\sigma}{\sqrt{n}}\right]$$

The upper-tail probability of $z = (1 - 0.8) / 2 = 0.1$

Using R-Commander, $z_{crit} = 1.28$

80% confidence interval estimation for μ is

$[110 - 1.28 \times 2, 110 + 1.28 \times 2] = [107.44, 112.56]$

Example 7

- For the question in Example 5, suppose that we did not know σ and estimated it using the sample standard deviation $s = 6$.
  a. Find the standard error for the sample mean as the estimator of the population mean.
  b. Find the 80% confidence interval estimation for μ based on this sample.

Answer 7

a. $SE = \frac{s}{\sqrt{n}} = \frac{6}{\sqrt{9}} = \frac{6}{3} = 2$

b. The confidence interval for the population mean at c confidence level is

$$\left[\bar{x} - t_{crit} \times \frac{s}{\sqrt{n}}, \bar{x} + t_{crit} \times \frac{s}{\sqrt{n}}\right]$$
$$= [110 - t_{crit} \times 2, 110 + t_{crit} \times 2]$$

We use the t-distribution with $9 - 1 = 8$ degrees of freedom.

t_{crit} : its upper tail probability is $(1 - 0.8) / 2 = 0.1$ (0.8 confidence)

Using R-Commander, $t_{crit} = 1.40$

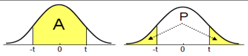
The 80% confidence interval estimation for μ based on this sample:

$[110 - 1.40 \times 2, 110 + 1.40 \times 2] = [107.2, 112.8]$

Table A-2 **The t-Table**

t-distribution showing area to the right

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
∞	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905



DF	A / P	0.80 / 0.20	0.90 / 0.10	0.95 / 0.05	0.98 / 0.02	0.99 / 0.01	0.995 / 0.005	0.998 / 0.002	0.999 / 0.001
1		3.078	6.314	12.706	31.820	63.657	127.321	318.309	636.619
2		1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3		1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5		1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		1.397	1.860	2.306	2.897	3.355	3.833	4.501	5.041
9		1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10		1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11		1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12		1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13		1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14		1.345	1.761	2.145	2.625	2.977	3.326	3.787	4.140
15		1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16		1.337	1.746	2.120	2.584	2.921	3.252	3.686	4.015
17		1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18		1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19		1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20		1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21		1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22		1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23		1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24		1.318	1.711	2.064	2.492	2.797	3.090	3.467	3.745
25		1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26		1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27		1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28		1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29		1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30		1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
31		1.309	1.695	2.040	2.453	2.744	3.022	3.375	3.633
32		1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622
33		1.308	1.692	2.035	2.445	2.733	3.008	3.356	3.611
34		1.307	1.691	2.032	2.441	2.728	3.002	3.348	3.601
35		1.306	1.690	2.030	2.438	2.724	2.996	3.340	3.591
36		1.306	1.688	2.028	2.434	2.719	2.991	3.333	3.582
37		1.305	1.687	2.026	2.431	2.715	2.985	3.326	3.574
38		1.304	1.686	2.024	2.429	2.712	2.980	3.319	3.566
39		1.304	1.685	2.023	2.426	2.708	2.976	3.313	3.558
40		1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
42		1.302	1.682	2.018	2.418	2.698	2.963	3.296	3.538
44		1.301	1.680	2.015	2.414	2.692	2.956	3.286	3.526
46		1.300	1.679	2.013	2.410	2.687	2.949	3.277	3.515
48		1.299	1.677	2.011	2.407	2.682	2.943	3.269	3.505
50		1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60		1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70		1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80		1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90		1.291	1.662	1.987	2.369	2.632	2.878	3.183	3.402
100		1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.391
120		1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
150		1.287	1.655	1.976	2.351	2.609	2.849	3.145	3.357
200		1.286	1.652	1.972	2.345	2.601	2.839	3.131	3.340
300		1.284	1.650	1.968	2.339	2.592	2.828	3.118	3.323
500		1.283	1.648	1.965	2.334	2.586	2.820	3.107	3.310
∞		1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Example 8

- A person visits her doctor with concerns about her blood pressure.
  - If the systolic blood pressure exceeds 150, the patient is considered to have high blood pressure and medication may be prescribed.
- A patient's blood pressure readings often have a considerable variation during a given day.
- Suppose a patient's systolic blood pressure readings during a given day have a normal distribution with a mean μ = 160 mm mercury and a standard deviation σ = 20 mm.

a. What is the probability that a single blood pressure measurement will fail to detect that the patient has high blood pressure?

b. If five blood pressure measurements are taken at various times during the day, what is the probability that the average of the five measurements will be less than 150 and hence fail to indicate that the patient has high blood pressure?

c. How many measurements would be required in a given day so that there is at most 1% probability of failing to detect that the patient has high blood pressure?

Answer 8

a. Let y be the blood pressure measurement of the patient.
  - y has a normal distribution with μ = 160 and σ = 20

    $P(_{\text{measurement fails to detect high pressure}}) = P(y \le 150) = P\left(z \le \frac{150-160}{20}\right) = P(z \le -0.5) = 0.3085$
  - Thus, there is over a 30% chance of failing to detect that the patient has high blood pressure if only a single measurement is taken.

b. Let $\bar{y}$ be the average blood pressure of the five measurements.
  - Then, $\bar{y}$ has a normal distribution with μ = 160 and σ = 20 $\quad \sigma = \frac{20}{\sqrt{5}} = 8.944$

Answer 8

$P(\bar{y} \le 150) = P\left(z \le \frac{150-160}{8.944}\right) = P(z \le -1.12) = 0.1314$
  - Therefore, by using the average of five measurements, the chance of failing to detect the patient has high blood pressure has been reduced from over 30% to about 13%.

c. We need to determine the sample size n such that $P(\bar{y} < 150) \le 0.01$.
  - Now, $P(\bar{y} < 150) = P\left(z \le \frac{150-160}{20/\sqrt{n}}\right)$.
  - From the normal tables, we have $P(z \le -2.326) = 0.01$.
    - Therefore, $\frac{150-160}{20/\sqrt{n}} = -2.326$.
    - Solving for n yealds n = 21.64.
  - It would require at least 22 measurements in order to achieve the goal of at most a 1% chance of failing to detect high blood pressure.

Example 9

- A company took a random sample of 30 first-year employees and asked them their level of satisfaction with their jobs.
  - It found that 80% of those sampled were "very happy" with their employment, ± 3% at a confidence level of 95%.
  - The company took this information and reported that 80% of all its employees were very happy with their jobs, ± 3%.
- Is that report is correct?

Answer 9

- Sample size n = 30, $\hat{p}$ is 0.8, and z^* is 1.96 (for %95 CI)
- Margin of error for a population proportion

$$MOE = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.8(1-0.8)}{30}} = 1.96\sqrt{\frac{0.8(0.2)}{30}}$$
$$= 1.96 \times 0.073 = 0.1431$$

- Convert the proportion to a percent by multiplying by 100%: 0.1431(100%) = 14.31%
  - This value is much larger than the reported MOE of 3%.

Example 10

- A poll of 1000 likely voters showed that Candidate Ali had 48% of the vote, and Candidate Veli had 52% of the vote.
- The margin of error was ± 3%, and the confidence level was 98%.
- Who is most likely to win the election?

31

Answer 10

- The margin of error is used to construct the confidence interval,
  - which is a range of likely values for the population parameter
    - here, the parameter is the percentage of all voters who would vote for a candidate.
- To calculate a confidence interval, you take the result from the sample and add and subtract the margin of error.
- In this case,
  - the 98% confidence interval for the proportion of all voters for Candidate Ali is 48% plus or minus 3%,
    - which is a range of 45% to 51%.

32

Answer 10

- For Candidate Veli,
  - the 98% confidence interval is 52% plus or minus 3%,
    - which is a range of 49% to 55%.
- Both confidence intervals contain possible values above 50%, so either candidate could win;
  - therefore, the results are too close to call.

33