MA331 Intermediate Statistics

Lecture 05 Inference on Population Mean and Proportion [1]

Xiaohu Li

Department of Mathematical Sciences
Stevens Institute of Technology
Hoboken, New Jersey 07030

Weeks 06-07

[1] Based on Chapters 6, 7 and 8.

1. Statistical hypothesis

✍ The sample mean $\bar{X}$ and the CI $\bar{X} \pm z_{1-\alpha/2}\sigma/\sqrt{n}$ directly estimate the unknown population mean μ.

✍ Sometimes, experimenter only wants to know the partial information of μ, e.g., $\mu > \mu_0$? — to test a statement about μ based on the sample.

✍ A statistical hypothesis is an assumption or a theory about the characteristics of population distributions, and the procedure to prove or disprove it based on the sample is called as a test of statistical significance.

✍ Weight of cherry tomato packs:

- Does the calibrating machine that sorts cherry tomatoes into packs need revision?
- Statistical hypothesis: Is the population mean μ for the weight of cherry tomato packages equal to the designed 227g (0.5pound)?

2. Null hypothesis and alternative hypothesis

✍ Null hypothesis (labeled H_0) is a very specific statement about the population parameter, usually it tends to be disproved.

✍ Alternative hypothesis (labeled H_a) is a more general statement about the population parameter, and H_0 and H_a are mutually exclusive.

✍ Null and Alternative are not of equal position. Researchers usually tend to protect H_0. Be careful to choose the null. E.g., FDA's attitude and manufacturer's attitude to a new drug.

✍ Weight of cherry tomato packages:

- H_0: $\mu = 227$ (the calibrating machine is good.)
- H_a: $\mu \neq 227$ (the machine needs to be revised.)

✍ For $H_0 : \mu = \mu_0$, if $H_a : \mu \neq \mu_0$ ($H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$), then the test is called two-sided (one-sided)/two-tailed (one-tailed).

3. Testing statistic

☞ To test the significance of H_0 we consider the sample $X_1, \cdots, X_n$, which is from the population X and hence carries the concerned information.

☞ To extract the useful information we resort to a statistic

$$T(X_1, \cdots, X_n),$$

some function of the sample and independent of unknown parameters.

☞ E.g., To test $H_0 : p_H = 1/2$ we use the sample proportion n_H/n with $n_H = (X_1 + \cdots + X_n)/n$.

QUERY 1: How do you define X_i for $i = 1, \cdots, n$?

☞ To test $H_0 : \mu = \mu_0$ (with known σ) we use

$$T(X_1, \cdots, X_n) = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \equiv Z.$$

4. p-value – evidence against the null hypothesis

☞ With the distribution of T under H_0, one may evaluate the chance of observing the data $x_1, \cdots, x_n$ at hand — p-value.

☞ A smaller (larger) p-value is against (in favor of) the null H_0.

☞ **Example**: The cherry tomato packaging process has a known $\sigma = 5$g.

- To test $H_0 : \mu = \mu_0 = 227$g versus $H_a : \mu \neq 227$g, a dozen randomly selected boxes are observed to have the average weight $\bar{x} = 222$g.
- Under H_0, $\bar{X} \sim \mathcal{N}(227, 5^2/12)$, and the prob of drawing a random sample with mean 5g (the observed distance) away from 227.

$$\begin{aligned} P(|\bar{X} - \mu_0| \geq |\bar{x} - \mu_0|) &= P\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \geq \frac{|222 - 227|}{5/\sqrt{12}}\right) \\ &= P\left(|Z| \geq \sqrt{12}\right) = 2\Phi(-2\sqrt{3}) \approx 0.0005. \end{aligned}$$

- The small p-value 0.0005 disfavors H_0. So, we reject $H_0 : \mu = 227$.

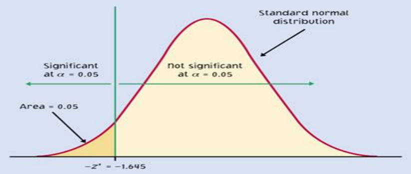
5. Significance level and critical values

✐ Usually, a p-value less than 0.05 is considered to be significant.

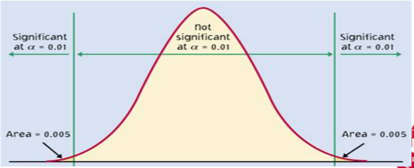
- The data observed is unlikely to be entirely due to randomness from the sampling. 0.1 and 0.01 are for looser and tougher alternatives, respectively.
- A significance level $\alpha = 0.1, 0.05, 0.01$ is always given in practice.

✐ Based on the distribution of T, the lower and upper $\alpha/2$ quantile $c_{\alpha/2}$ and $c_{1-\alpha/2}$ serve as the critical values.

✐ Since $T \notin (c_{\alpha/2}, c_{1-\alpha/2})$ is equivalent to a p-value smaller than α, we can also make the decision by comparing the observed T with critical values.



(a) p-value 0.05

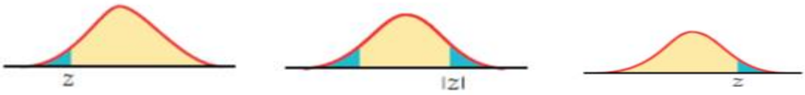


(b) p-value 0.01

6. z-test for a population mean μ

At significance level α, test $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ based on $(x_1, \cdots, x_n)$ from pop $X \sim \mathcal{N}(\mu, \sigma^2)$ with known σ.

- Under H_0, test statistic $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$, observed as $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
- p-value method: reject H_0 if $\mathrm{P}(|Z| \geq |z|) = 2\Phi(-|z|) < \alpha$.
- Critical value method: reject H_0 if the observed $z \notin (z_{\alpha/2}, z_{1-\alpha/2})$, where $z_{\alpha/2}$, $z_{1-\alpha/2}$ are lower and upper $\alpha/2$ quantiles of $\mathcal{N}(0,1)$.
- One side test: reject H_0 if $\begin{cases} \mathrm{P}(Z > z) < \alpha \quad \text{or} \quad z > z_{1-\alpha}, & \text{for } H_a : \mu > \mu_0, \\ \mathrm{P}(Z < z) < \alpha \quad \text{or} \quad z < z_{\alpha}, & \text{for } H_a : \mu < \mu_0. \end{cases}$
- Rejection regions



(c) $H_a : \mu < \mu_0$ (d) $H_a : \mu \neq \mu_0$ (e) $H_a : \mu > \mu_0$

7. A revisit to the cherry tomato packaging process

With a known standard deviation $\sigma = 5$g, test, at the significance level $\alpha = 0.05$,

$$H_0 : \mu = \mu_0 = 227 \qquad \text{versus} \qquad H_a : \mu \neq 227$$

based on a randomly selected box with average weight $\bar{x} = 222$g.

- Under H_0, the testing statistic $Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim \mathcal{N}(0, 1)$.
- Based on the sample, Z is observed as

$$z = \frac{222 - 227}{5/\sqrt{12}} = -\sqrt{12}.$$

  Note that the p-value

$$\mathrm{P}\left(|Z| \geq |z| = \sqrt{12}\right) \approx 0.0005 << 0.05 = \alpha.$$

  We reject $H_0 : \mu = 227$. That is, the machine does need a calibration.

- Critical values $z_{.025} = -1.96$, $z_{.975} = 1.96$. Since

$$z = -3.46 \notin (-1.96, 1.96),$$

  we reject $H_0 : \mu = 227$ again.

8. Use and abuse of the test for significance

✐ In statistics $\alpha = 0.05$ is well accepted. A standard significance level is typically preferred in a specific area.

- a smaller α may be appropriate if rejecting H_0 has 'costly' implications, e.g., global warming, convicting with DNA evidence.
- a larger α makes it less likely to miss an interesting result in a preliminary study.
- No sharp 'cutoffs', e.g., 4.9% versus 5.1%. Usually, we describe the evidence by p-value itself, and the magnitude matters: significant, somewhat/very significant.

✐ Statistical significance only logically tells whether the effect observed is likely to be due to chance (randomness) alone. Without the magnitude of the effect it may be of no practical interest.

✐ E.g., a new drug lowers patient temperature by 0.4° (p-value < 0.01). But clinical benefit of temperature reduction only appears for a decrease no less than 1°.

9. Lack of significance and effect

✐ Indeed, failing to find statistical significance in experimental results means

not rejecting H_0, however this doesn't imply to accept it.

The sample size, for instance, could be too small to overcome large variability in the population.

✐ No consensus on how big an effect is to be considered meaningful. Sometimes, effects seeming trivial can be very important in reality.

✐ E.g., improving the format of a computerized test reduces the average response time by 2 seconds, a small effect; however, it is done millions of times a year, and the cumulated time savings will be gigantic.

✐ Always think about the context. Try to plot your results, and compare them with a baseline or results from similar studies.

10. Two types of errors

The statistical hypothesis is tested based on p-value — the logical evidence, two types of error may occur when a decision is made.

✐ Type I error

- Reject H_0 and H_0 is actually true (deny a true fact).
- Probability of making a type I error is controlled by the significance level α.

✐ Type II error

- Fail to reject actually false H_0 (fail to deny a false fact).
- Probability of making a type II error is β.
- $1-\beta$, the probability to reject a false H_0, is called the power of the test.

	H_0 true	H_a true
Reject H_0	Type I error	Correct decision
Accept H_0	Correct decision	Type II error

✐ α and β can not be simultaneously controlled smaller enough, and we set a small α only to protect H_0 and thus the stance of the researcher matters.

Query 2: If $\alpha+\beta=1$? Why?

11. Student t distribution

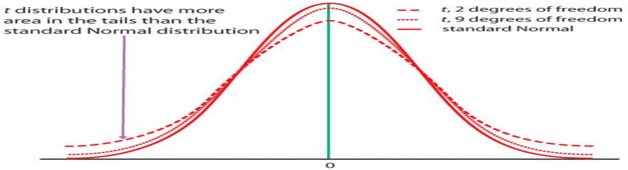
At significance level α, $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ based on $(X_1, \cdots, X_n) = (x_1, \cdots, x_n)$ from $X \sim \mathcal{N}(\mu, \sigma^2)$ with unknown σ.

✐ z-test fails because $Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}$ is not accessible any more.

✐ An estimation of σ is the sample standard error $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$. Naturally we turn to $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S}$.

✐ z-test evaluates p-value based on $Z \sim \mathcal{N}(0, 1)$. Likewise, the null distribution of T is necessary to get p-value in this context.

✐ W. S. Gosset (1908): Under $H_0 : \mu = \mu_0$, the above T has Student's t distribution with degree of freedom (df) $n - 1$.



12. t-test for a population mean μ

At significance level α, test $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ based on $(x_1, \cdots, x_n)$ from $X \sim \mathcal{N}(\mu, \sigma^2)$ with unknown σ.

✐ Under H_0, the testing statistic $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$, observed as $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$.

✐ p-value method: reject H_0 if the p-value $\mathrm{P}(|T| \geq |t|) < \alpha$.

✐ Critical value method: reject H_0 if $t \notin (t_{\alpha/2}(n-1), t_{1-\alpha/2}(n-1))$, where $t_{\alpha/2}(n-1)$ and $t_{1-\alpha/2}(n-1)$ are $\alpha/2$ and $(1-\alpha/2)$ quantiles respectively.

✐ One side test: reject H_0 if

$$\begin{cases} \mathrm{P}(T > t) < \alpha \quad \text{or} \quad t > t_{1-\alpha}(n-1), & \text{for } H_a : \mu > \mu_0, \\ \mathrm{P}(T < t) < \alpha \quad \text{or} \quad t < t_{\alpha}(n-1), & \text{for } H_a : \mu < \mu_0. \end{cases}$$

✐ Rejection regions

$$\begin{cases} (-\infty, t_\alpha(n-1)), & \text{for } H_a : \mu < \mu_0, \\ (-\infty, t_{\alpha/2}(n-1)) \cup (t_{1-\alpha/2}(n-1), \infty), & \text{for } H_a : \mu \neq \mu_0, \\ (t_{1-\alpha}(n-1), \infty), & \text{for } H_a : \mu > \mu_0. \end{cases}$$

13. z-table and t-table

✐ z-test gets p-value by standard normal density curve

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

✐ t-test evaluates p-value by Student's t density curve with df n

$$f_n(x) = \frac{\Gamma(n/2)}{\sqrt{(n-1)\pi}\Gamma((n-1)/2)}\left(1 + \frac{x^2}{n-1}\right)^{n/2}.$$

✐ p-value is concerned with the tail area,

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)\mathrm{d}x \qquad \text{and} \qquad \int_{-\infty}^{t} f_{n-1}(x)\mathrm{d}x$$

are both not solvable. So, we resort to z table and t table.

✐ In R: we do it by pnorm(z,0,1) and pt(t,n-1), resp.

QUERY 3: How do you obtain upper tail area $\int_z^\infty \phi(x)\mathrm{d}x$ and $\int_t^\infty f(x)\mathrm{d}x$?

14. z-table and t-table

TABLE A Standard normal probabilities

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

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

15. A revisit to the cherry tomato packaging process

✌ At the significance level $\alpha = 0.05$ and without the knowledge of σ, test

$$H_0 : \mu = \mu_0 = 227 \qquad \text{versus} \qquad H_a : \mu < 227$$

based on a randomly selected box with average weight $\bar{x} = 222$g and standard error $s = 6$.

✌ Under H_0, $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{s} \sim t(12-1)$. Based on the sample values, it evaluates that $t = \sqrt{12}(222-227)/6 = -2.887$.

- p-value

  $$\mathrm{P}(T < -2.887) \approx 0.007 << 0.05 = \alpha.$$

  Then we reject $H_0 : \mu = 227$. That is, the machine does need a calibration.

- Critical values $t_{0.05}(11) = -1.706$, and due to the observed $t = -2.887 < -1.706$ we reject $H_0 : \mu = 227$ again.

16. Summary of inference on a single population mean

- Given a confidence level $1-\alpha$, the CI
  - with a known σ: $\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$.
  - with σ unknown: $\bar{x} \pm t_{1-\alpha/2}(n-1) \frac{s}{\sqrt{n}}$.
- A test of hypotheses is accomplished through the following steps:
  1. State the null hypotheses H_0 and select the alternative H_a (one side or two sides).
  2. Choose a significance level α.
  3. Calculate the observed testing statistics
     - with a known σ: $z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}$.
     - with σ unknown: $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s}$ and d.f. $n-1$.
  4. Find the tail area under the z-curve and t-curve, respectively.
  5. State the p-value, draw a conclusion and interpret your result.
- In R, we use ztest(sample) and ttest(sample) to perform calculation, respectively.

17. Matched pairs t-test procedures

✍ Two samples of matched observations are recorded to compare treatments/control at the individual level. For examples,

- Engineering: in pre-test and post-test studies, data collected on the same sample elements before and after an experiment.
- Biometrics: twin studies tries to sort out the influence of genetic factors by comparing a variable between sets of twins.
- Social science: using people matched for age, sex, and education to screen out the effect due to lurking factors.

✍ In these cases, we use the paired data to test the difference between two population means. The variable here becomes the difference $D = X - Y$, and hypotheses are then

$$H_0 : \mu_D = \mu_X - \mu_Y = \mu_0 \quad \text{versus} \quad H_a : \mu_D < \mu_0.$$

Technically, this is of no difference from the tests for a single population.

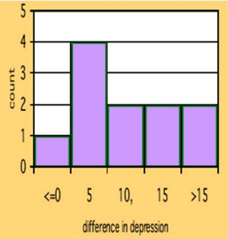
18. An application: Lack of caffeine increase depression?

✌ Individuals diagnosed as caffeine-dependent are deprived of caffeine-rich foods and assigned to receive daily pills. Sometimes, the pills contain caffeine and other times they contain a placebo. Depression was assessed in both situations.

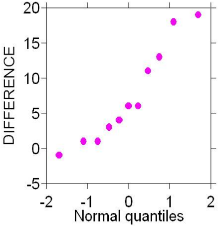
✌ Each subject gets 2 variables, but well only look at the difference. The sample distribution appears appropriate for a t-test.

Subject	Depression with Caffeine	Depression with Placebo	Placebo - Cafeine
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	-1

(f) paired data



11 “difference” data points.



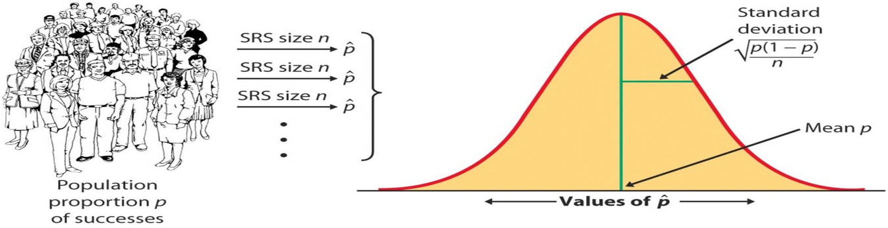
(g) Normal quantile plot

19. Lack of caffeine increase depression? continued

- Based on $d_1, \cdots, d_{11}$, we evaluate $\bar{d} = 7.38$ and $s = 6.92$ with $n = 11$.
- $H_0 : \mu_D = 0$ versus $H_a : \mu_D > 0$ at significance level $\alpha = 0.05$.
- T is observed as $t = \sqrt{n}(\bar{d} - \mu_0)/s = 3.53$.
  - Since p-value $P(T > t = 3.53) \approx 0.0027 << 0.05 = \alpha$, we reject $H_0 : \mu_D = 0$, i.e., caffeine does matters.
  - Critical values $t_{1-0.05}(10) = 1.81$, and due to $t = 3.53 > 1.81$ we reject $H_0 : \mu_D = 0$ again.
- Some important factors:
  - The smaller sample size $n = 11$ fails to carry enough information. For $n < 15$, the data must be close to normal and without outliers, for $15 > n > 40$, mild skewness is acceptable but not outliers, and it is valid even with strong skewness for $n > 40$.
  - The sample's deviation from normality undermines the t-test. So, check it using QQ plot. For non-normal data, resort to transformation and nonparametric methods.

20. Distribution of sample proportion

✌ Based on CLT, the distribution of a sample proportion $\hat{p}$ is approximately normal, i.e., $\hat{p} \sim \mathcal{N}(p, p(1-p)/n)$, when the sample size is large enough.



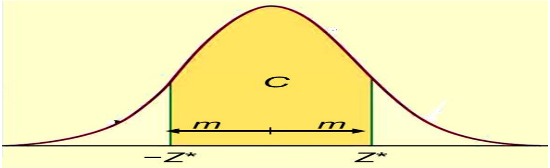
✌ The conditions for inference on the population proportion p:

- Data used for the estimate $\hat{p}$ are an SRS from the population to be studied.
- Sample size n is large enough so that the sampling distribution can be approximated with a normal distribution.

21. Large-sample confidence interval for p

✌ For an SRS of size n from a large population and with sample proportion $\hat{p}$ calculated from the data, an approximate level $1 - \alpha$ confidence interval for p is

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$



- Margin of error $m = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, the quantile $z^* = z_{1-\alpha/2}$.
- $c = 1 - \alpha$ is the area under the standard normal curve between $\pm z^*$.

✌ This method is suitable when the number of successes and the number of failures are both at least 15.

22. Test for a population proportion p

At significance level α, test $H_0 : p = p_0$ vs $H_a : p \neq p_0$ based on $(x_1, \cdots, x_n)$ from population X with probability of success p.

✐ Under H_0, the testing statistic $Z \sim \mathcal{N}(0, 1)$, which is observed based on the sample as

$$z = \frac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1 - p_0)}}.$$

✐ p-value method: reject H_0 if the p-value

$$\mathrm{P}(|Z| \geq |z|) \approx 2\Phi(-|z|) < \alpha.$$

✐ Critical value method: reject H_0 if the observed $z \notin (z_{\alpha/2}, z_{1-\alpha/2})$, where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are lower and upper $\alpha/2$ quantiles of $\mathcal{N}(0, 1)$.

✐ One-tail tests:

- For $H_a : p > p_0$, reject H_0 if $z > z_{1-\alpha}$ or p-value $\mathrm{P}(Z > |z|) < \alpha$.
- For $H_a : p < p_0$, reject H_0 if $z < z_\alpha$ or p-value $\mathrm{P}(Z < -|z|) < \alpha$.