HYPOTHESIS TESTING

Null hypothesis : H_0 : $\mu = 20$

Alternate Hypothesis : H_1 : $\mu \neq 20$ or $\mu > 20$ or $\mu < 20$

level of significance $= \alpha : 0.05$ or 0.01 or 0.10

α = P [rejecting H_0 when H_0 is true] = P [Type - I Error]

β = Accepting false H_0 = Type II error

4 types of decisions

- In the hypothesis-testing situation, there are four possible outcomes.

	Ho True	Ho False
Reject Ho	Error (Type – I)	Correct Decision
Do not reject Ho	Correct Decision	Error (Type – II)

P(Type-I) = Level of significance

P(Type-II) = β

Statistical Tests

↳ z - test

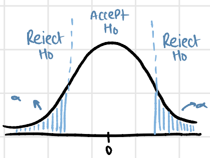
↳ t - test

↳ f - test

Two Tailed Test

$H_0 : \mu = K$

$H_1 : \mu \neq K$



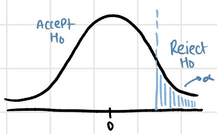
$$\alpha = \frac{1 - \text{confidence level}}{2}$$

One Tailed Test

Right tailed test

$H_0 : \mu = K$

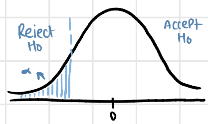
$H_1 : \mu > K$



Left-tailed test

$H_0 : \mu = K$

$H_1 : \mu < K$



$$\alpha = 1 - \text{confidence level}$$

Z-Test for mean

↳ $n \geq 30$

↳ σ is known

↳ independent sample

↳ Population is normally distributed

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

sample mean → $\bar{x}$; H_0 mean → μ; standard deviation → σ; sample size → n

$n \geq 30$

1. State H_0, H_1
2. calculated z formula
3. find rejection region
4. If calculated z inside rejection regeion
   ↳ reject H_0

1. A factory has a machine that dispenses 80 mL of fluid in a bottle. An employee believes the average amount of fluid is not 80 mL. Using 40 samples, he measures the average amount dispensed by the machine to be 78 mL with a standard deviation of 2.5. (a) State the null and alternative hypotheses. (b) At a 95% confidence level, is there enough evidence to support the idea that the machine is not working properly?

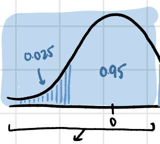
$\tilde{x} = 78$, $S = 2.5$, $n = 40$ → as >30 so z-test

① null hypotheses: $H_0 = \mu = 80$ alternate hypotheses: $H_a = \mu \neq 80$

② $z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{78 - 80}{2.5 / \sqrt{40}} = -5.06$

③ $\alpha = \frac{1 - 0.95}{2} = 0.025$

0.005 to 0.99



afrea under left curve

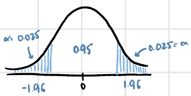
A_L = confidence level + α

$A_L = 0.95 + 0.025 = 0.975$ → find on Table

$Z_{\frac{\alpha}{2}} = 1.96$

④ Z falls in rejection region

Reject H_0



So with 95% confidence we belive that machine is not working properly

Question # 1 A random sample of size 36 is taken from a normal population with a known variances $\sigma^2 = 25$. If the mean of the sample is $\bar{X} = 42.6$, test the null hypothesis $\mu = 45$ against the alternativ hypothesis $\mu < 45$ with $\alpha = 0.05$.

$n > 30$ so z test

$n = 36$, $\sigma^2 = 25$, $\bar{x} = 42.6$, $\alpha = 0.05$

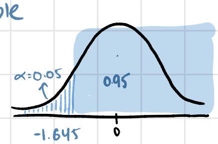
① $H_0 = 45$, $H_1 < 45$, $\sigma = 5$

② $z = \frac{42.6 - 45}{5 / \sqrt{36}} = -2.88$

③ area under right curve

$A_R = 0.95$ → check table

$Z_{\frac{\alpha}{2}} = -1.645$



④ Z falls in rejection region

Reject H_0

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

→ Closest to 0.95
→ take average
→ $\frac{0.04 + 0.05}{2} = 0.045$
$= 1.645$

Z-test for two means

- ↳ $n > 30$
- ↳ σ known
- ↳ independent sample

sample mean H₀ mean

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standard deviation

sample size

$$\delta = \sqrt{\frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}}$$

An admission test was administered to incoming freshmen in the College of Medical Laboratory and Sciences and College of Radiologic Technology with 100 students each college randomly selected. The mean scores of the given samples were $\bar{x}_1 = 90$ and $\bar{x}_2 = 85$ and the variances of the test scores were 40 and 35 respectively. Is there a significant difference between the two groups? Use .01 level of significance.

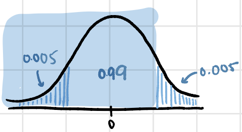
$\bar{X}_1 = 90$ $\bar{X}_2 = 85$ $n = 100$

$s^2 = 40$ $s^2 = 35$ $\alpha = 0.01$

① $H_0: \mu = 0$

$H_1: \mu \neq 0$

② $$z = \frac{(90 - 85) - (0 - 0)}{\sqrt{\frac{40^2}{100} + \frac{35^2}{100}}} = 5.774$$

③ 

$AL = 0.005 + 0.99 = 0.995$

$Z_{\frac{\alpha}{2}} = 2.575$

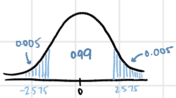
$= \frac{0.07 + 0.08}{2} = 0.075$

$= 2.575$

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

④ z falls in rejected region

H₀ rejected



T-Test for mean

- ↳ variance > 1
- ↳ $n < 30$
- ↳ σ is known
- ↳ independent sample
- ↳ Population is normally distributed

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

sample mean ($\bar{x}$), H_0 mean (μ), standard deviation (σ), sample size (n)

$df = n - 1$ ↳ degree of freedom

1. State H_0, H_1 $n<30$
2. calculated t formula
3. find rejection region
4. If calculated t inside rejection region
   ↳ reject H_0

2. A company manufactures car batteries with an average life span of 2 or more years. An engineer believes this value to be less. Using 10 samples, he measures the average life span to be 1.8 years with a standard deviation of 0.15. (a) State the null and alternative hypotheses. (b) At a 99% confidence level, is there enough evidence to discard the null hypothesis?

① $H_0 = \mu \geq 2$

$H_a : \mu < 2$

②

$\mu_0 = 2,\ \bar{x} = 1.8,\ s = 0.15,\ n = 10$ → as $n<30$ so we use t test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.8 - 2}{0.15/\sqrt{10}} = -4.22$$

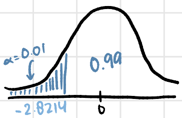
③ $df = n - 1$

$= 10 - 1 = 9$

$\alpha = 1 - 0.99 = 0.01$

find on table

$t_{\frac{\alpha}{2}} = -2.821$



④ t falls in rejection region

H_0 rejected

so with 99% confidence we belive that H_a

df \ α	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	318.3
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	22.33
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	10.21
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.787

T-test for two means

↳ $n < 30$

↳ σ known

↳ independent samples

Kaito grows tomatoes in two separate fields. When the tomatoes are ready to be picked, he is curious as to whether the sizes of his tomato plants differ between the two fields. He takes a random sample of plants from each field and measures the heights of the plants. Here is a summary of the results:

	Field A	Field B
Mean	1.3 m	1.6 m
Standard deviation	0.5 m	0.3 m
Number of plants	22	24

↳ $\sigma_1 \neq \sigma_2$ variance

sample mean H₀ mean

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standard deviation

sample size

$$df = \frac{(s_1 + s_2)^2}{\frac{s_1^2}{n-1} + \frac{s_2^2}{n-1}}$$

↳ $\sigma_1 = \sigma_2$ variance

$$s = \frac{(n-1)(s_1)^2 + (n-1)(s_2)^2}{(n+n)-2}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

① $H_0 : \mu = 0 \rightarrow \mu_1 = \mu_2$

$H_1 : \mu \neq 0 \rightarrow \mu_1 \neq \mu_2$

assume (as not mentioned)

② $n_1 = 22$ $\quad \sigma_1 = 0.5$ $\quad \bar{x}_1 = 1.3$ $\quad \alpha = 0.05$

$n_2 = 24$ $\quad \sigma_2 = 0.3$ $\quad \bar{x}_2 = 1.6$

$$t = \frac{(1.3 - 1.6)(0 - 0)}{\sqrt{\frac{0.5^2}{22} + \frac{0.3^2}{24}}} = -2.4402$$

③ $df_1 = 21$ $\quad df_2 = 23$ $\quad \alpha = 0.05/2 = 0.025$

take $t_{\frac{\alpha}{2}}$ for the smallest df

$t_{,\frac{\alpha}{2}} = 2.080$

df	0.20	0.10	0.05	0.04	0.03	0.025	0.02	0.01	0.005	0.0005
1	1.376	3.078	6.314	7.916	10.579	12.706	15.895	31.821	63.657	636.619
2	1.061	1.886	2.920	3.320	3.896	4.303	4.849	6.965	9.925	31.599
3	0.978	1.638	2.353	2.605	2.951	3.182	3.482	4.541	5.841	12.924
4	0.941	1.533	2.132	2.333	2.601	2.776	2.999	3.747	4.604	8.610
5	0.920	1.476	2.015	2.191	2.422	2.571	2.757	3.365	4.032	6.869
6	0.906	1.440	1.943	2.104	2.313	2.447	2.612	3.143	3.707	5.959
7	0.896	1.415	1.895	2.046	2.241	2.365	2.517	2.998	3.499	5.408
8	0.889	1.397	1.860	2.004	2.189	2.306	2.449	2.896	3.355	5.041
9	0.883	1.383	1.833	1.973	2.150	2.262	2.398	2.821	3.250	4.781
10	0.879	1.372	1.812	1.948	2.120	2.228	2.359	2.764	3.169	4.587
11	0.876	1.363	1.796	1.928	2.096	2.201	2.328	2.718	3.106	4.437
12	0.873	1.356	1.782	1.912	2.076	2.179	2.303	2.681	3.055	4.318
13	0.870	1.350	1.771	1.899	2.060	2.160	2.282	2.650	3.012	4.221
14	0.868	1.345	1.761	1.888	2.046	2.145	2.264	2.624	2.977	4.140
15	0.866	1.341	1.753	1.878	2.034	2.131	2.249	2.602	2.947	4.073
16	0.865	1.337	1.746	1.869	2.024	2.120	2.235	2.583	2.921	4.015
17	0.863	1.333	1.740	1.862	2.015	2.110	2.224	2.567	2.898	3.965
18	0.862	1.330	1.734	1.855	2.007	2.101	2.214	2.552	2.878	3.922
19	0.861	1.328	1.729	1.850	2.000	2.093	2.205	2.539	2.861	3.883
20	0.860	1.325	1.725	1.844	1.994	2.086	2.197	2.528	2.845	3.850
21	0.859	1.323	1.721	1.840	1.988	2.080	2.189	2.518	2.831	3.819
22	0.858	1.321	1.717	1.835	1.983	2.074	2.183	2.508	2.819	3.792
23	0.858	1.319	1.714	1.832	1.978	2.069	2.177	2.500	2.807	3.768
24	0.857	1.318	1.711	1.828	1.974	2.064	2.172	2.492	2.797	3.745
25	0.856	1.316	1.708	1.825	1.970	2.060	2.167	2.485	2.787	3.725
26	0.856	1.315	1.706	1.822	1.967	2.056	2.162	2.479	2.779	3.707
27	0.855	1.314	1.703	1.819	1.963	2.052	2.158	2.473	2.771	3.690
28	0.855	1.313	1.701	1.817	1.960	2.048	2.154	2.467	2.763	3.674
29	0.854	1.311	1.699	1.814	1.957	2.045	2.150	2.462	2.756	3.659
30	0.854	1.310	1.697	1.812	1.955	2.042	2.147	2.457	2.750	3.646
40	0.851	1.303	1.684	1.796	1.936	2.021	2.123	2.423	2.704	3.551
50	0.849	1.299	1.676	1.787	1.924	2.009	2.109	2.403	2.678	3.496
60	0.848	1.296	1.671	1.781	1.917	2.000	2.099	2.390	2.660	3.460
70	0.847	1.294	1.667	1.776	1.912	1.994	2.093	2.381	2.648	3.435
80	0.846	1.292	1.664	1.773	1.908	1.990	2.088	2.374	2.639	3.416
100	0.845	1.290	1.660	1.769	1.902	1.984	2.081	2.364	2.626	3.390
140	0.844	1.288	1.656	1.763	1.896	1.977	2.073	2.353	2.611	3.361
1000	0.842	1.282	1.646	1.752	1.883	1.962	2.056	2.330	2.581	3.300
∞	0.842	1.282	1.645	1.751	1.881	1.960	2.054	2.326	2.576	3.291
	60%	80%	90%	92%	94%	95%	96%	98%	99%	99.9%

CRITICAL VALUE FOR CONFIDENCE LEVEL

t lies in rejection region

H_0 rejected

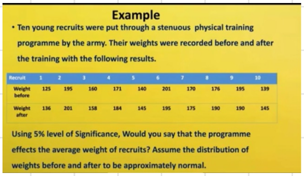
T-test for two means

↳ $n < 30$

↳ σ unknown

↳ dependent samples

1. $D = x_1 - x_2$
2. $D^2 = (x_1 - x_2)^2$
3. $\bar{X} = \frac{\Sigma D}{n}$
4. $S.d = \sqrt{\frac{n \Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$
5. $t = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$
6. $df = n - 1$



Example

- Ten young recruits were put through a stenuous physical training programme by the army. Their weights were recorded before and after the training with the following results.

Recruit	1	2	3	4	5	6	7	8	9	10
Weight before	125	195	160	171	140	201	170	176	195	139
Weight after	136	201	158	184	145	195	175	190	190	145

Using 5% level of Significance, Would you say that the programme effects the average weight of recruits? Assume the distribution of weights before and after to be approximately normal.

↳ $\sigma_1 = \sigma_2$ variance

$$\delta = \frac{(n-1)(\delta_1)^2 + (n-1)(\delta_2)^2}{(n+n) - 2}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\delta \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\delta = \sqrt{\frac{n \Sigma x^2 - (\Sigma x)^2}{n(n-1)}}$$

① $H_0 : \mu = 0 \qquad H_1 : \mu \neq 0$

② $n = 10$, $\alpha = 0.05$

$$\Sigma D = (125-136) + (195-201) + (160-158) + (171-184) + (140-145) + (201-195) + (170-175) + (176+190) + (195-190) + (139-145) = -47$$

$$\bar{X} = \frac{47}{10} = 4.7$$

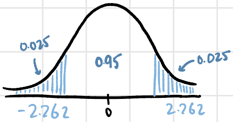
$$\Sigma D^2 = (125-136)^2 + (195-201)^2 + (160-158)^2 + (171-184)^2 + (140-145)^2 + (201-195)^2 + (170-175)^2 + (176+190)^2 + (195-190)^2 + (139-145)^2 = 673$$

$$S.d = \sqrt{\frac{10(673) - (47)^2}{10(10-1)}} = 7.09$$

③ $t = \frac{4.7 - 0}{7.09 / \sqrt{10}} = 2.09$

④ $df = 9$, $\alpha = \frac{0.05}{2} = 0.025$

$t_{\frac{\alpha}{2}} = 2.262$



df \ α	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	318.3
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	22.33
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	10.21
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.787

⑤ t lies in acceptence region

Ho accept

CHP 10

10.20 A random sample of 64 bags of white cheddar popcorn weighed, on average, 5.23 ounces with a standard deviation of 0.24 ounce. Test the hypothesis that $\mu = 5.5$ ounces against the alternative hypothesis, $\mu < 5.5$ ounces, at the 0.05 level of significance.

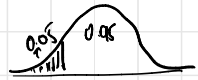
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.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

→ Z test

$n = 64$ $\quad \bar{X} = 5.23$ $\quad \delta = 0.24$ $\quad \alpha = 0.05$

$H_0: \mu = 5.5$

$H_1: \mu < 5.5$



-1.645

$$z = \frac{5.23 - 5.5}{0.24/\sqrt{64}} = -9$$

$A_R = 0.95$

from table $= 1.645$

$P = -1.645$

Ho reject $\mu < 5.5$

10.23 Test the hypothesis that the average content of containers of a particular lubricant is 10 liters if the contents of a random sample of 10 containers are 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, and 9.8 liters. Use a 0.01 level of significance and assume that the distribution of contents is normal.

$n = 10$ → T test $\quad \alpha = 0.01$

10.3 10.4 10.2 10.1 10.1 9.9 10.3 9.8 9.7 9.8

$H_0: \mu = 10$

$H_1: \mu \neq 10$

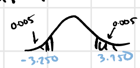
$$\bar{X} = \frac{100.6}{10} = 10.06$$

$\Sigma X^2 = 1012.64$

$$\delta = \sqrt{\frac{1012.64}{10} - \left(\frac{100.6}{10}\right)^2} = 0.246$$

$$t = \frac{10.06 - 10}{0.246/\sqrt{10}} = 0.77$$

df 9



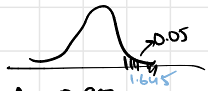
$P = 3.250$

df \ α	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	318.3
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	22.33
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	10.21
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.787

Ho accept

10.25 It is claimed that automobiles are driven on average more than 20,000 kilometers per year. To test this claim, 100 randomly selected automobile owners are asked to keep a record of the kilometers they travel. Would you agree with this claim if the random sample showed an average of 23,500 kilometers and a standard deviation of 3900 kilometers? Use a P-value in your conclusion.

H_1



$A_L = 0.95$

$P = 1.645$

z test

$H_0 \to \mu = 20{,}000$ $\quad n = 100 \quad \bar{x} = 23500 \quad \delta = 3900$

$H_1 \to \mu > 20{,}000$ $\quad \alpha = 0.05$ as not mentioned

H_0 reject

$\mu > 20{,}000$

$$z = \frac{23500 - 20000}{3900/\sqrt{100}} = 8.974$$

10.27 A study at the University of Colorado at Boulder shows that running increases the percent resting metabolic rate (RMR) in older women. The average RMR of 30 elderly women runners was 34.0% higher than the average RMR of 30 sedentary elderly women, and the standard deviations were reported to be 10.5 and 10.2%, respectively. Was there a significant increase in RMR of the women runners over the sedentary women? Assume the populations to be approximately normally distributed with equal variances. Use a P-value in your conclusions.

Decision: Reject H_0 and claim $\mu > 220$ milligrams.

10.27 The hypotheses are

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 > \mu_2.$$

Since $s_p = \sqrt{\frac{(29)(10.5)^2+(29)(10.2)^2}{58}} = 10.35$, then

$$P\left[T > \frac{34.0}{10.35\sqrt{1/30+1/30}}\right] = P(Z > 12.72) \approx 0.$$

Hence, the conclusion is that running increases the mean RMR in older women.

z test $\quad \delta$ known

$H_0 \to \mu_1 = \mu_2$ $\quad n = 30 \quad \bar{x}_1 - \bar{x}_2 = 34 \quad \delta_1 = 10.5$

$H_1 \to \mu_1 \neq \mu_2$ $\quad \delta = 10.2$

$$\delta = \frac{(29)(10.5)^2 - (29)(10.2)^2}{30+30-2} = 10.35$$

$$z = \frac{34}{10.35/\sqrt{\frac{1}{30}+\frac{1}{30}}} = 0.8481$$

z table

$\alpha = 0.05$

$A_L = 0.975$

$P = 1.96$

H_0 accept

10.35 To find out whether a new serum will arrest leukemia, 9 mice, all with an advanced stage of the disease, are selected. Five mice receive the treatment and 4 do not. Survival times, in years, from the time the experiment commenced are as follows:

Treatment	2.1	5.3	1.4	4.6	0.9
No Treatment	1.9	0.5	2.8	3.1	

At the 0.05 level of significance, can the serum be said to be effective? Assume the two populations to be normally distributed with equal variances.

T test δ unknown

$n_1 = 5$ $\bar{x}_1 = \frac{14.3}{5} = 2.86$ $\delta_1 = \frac{5(56.43)-(14.3)^2}{5(4)} = 3.883$

$n_2 = 4$ $\bar{x}_2 = \frac{8.3}{4} = 2.075$ $\delta_2 = \frac{4(21.31\ )-(8.3)^2}{4(3)} = 1.3625$

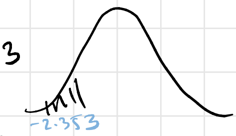
$x_1 - x_2 = 0.785$

$$S = \frac{3(1.3625)^2 + 4(3.883)^2}{5+4-2} = 1.674$$

$df = 3$

$\alpha = 0.05$

$P = 2.353$



$$t = \frac{0.785}{1.674/\sqrt{\frac{1}{5}+\frac{1}{4}}} = -0.70$$

$H_0 \rightarrow \mu_1 - \mu_2 = 0$

$H_1 \rightarrow \mu_1 - \mu_2 < 0$

H_0 accept

normal with equal variances.

10.39 The following data represent the running times of films produced by two motion-picture companies:

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Test the hypothesis that the average running time of films produced by company 2 exceeds the average running time of films produced by company 1 by 10 minutes against the one-sided alternative that the difference is less than 10 minutes. Use a 0.1 level of significance and assume the distributions of times to be approximately normal with unequal variances.

10.39 The hypotheses are

$x_1 - x_2 = 10$ $n_1 = 5$ $n_2 = 7$ $\alpha = 0.1$

$$H_0: \mu_{II} - \mu_I = 10,$$
$$H_1: \mu_{II} - \mu_I > 10.$$

$\alpha = 0.1$.
Degrees of freedom is calculated as

$$v = \frac{(78.8/5 + 913.333/7)^2}{(78.8/5)^2/4 + (913/333/7)^2/6} = 7.38,$$

hence we use 7 degrees of freedom with the critical region $t > 1.415$.
Computation: $t = \frac{(110-97.4)-10}{\sqrt{78.800/5+913.333/7}} = 0.22$.
Decision: Fail to reject H_0.

ANALYSIS OF VARIANCE (ANOVA)

F test

- ↳ independent samples
- ↳ normal distribution
- ↳ σ are same

not t-test as the more means, the more t-tests needed

1. State Ho, Ha, α
2. Degree's of freedom
   - ↳ df between groups = $K-1$
   - ↳ df within groups = $N-K$ → no of groups
     - ↳ total no of observations
   - ↳ diff total = $N-1$
   - ↳ F critical → using table
3. Sum of square Deviations from mean
   - ↳ each groups mean → $\bar{x}_1, \bar{x}_2, \bar{x}_3$
   - ↳ Grand mean → $\bar{\bar{x}}$
   - ↳ $SS_{total} = \Sigma(x-\bar{\bar{x}})^2$
   - ↳ $SS_{within} = \Sigma(x-\bar{x})^2$
   - ↳ $SS_{between} = SS_{total} - SS_{within}$
4. Mean Square
   - ↳ $MS_{between} = \frac{SS_{between}}{df_{between}}$
   - ↳ $MS_{within} = \frac{SS_{within}}{df_{within}}$
5. Calculate F statistic
   - ↳ $F = \frac{MS_{between}}{MS_{within}}$
   - ↳ conclusion

reject Ho	accept Ho
$F_{ratio} > F_{critical}$	$F_{ratio} < F_{critical}$
$F_{ratio} > 1$	$F_{ratio} \leq 1$

ANALYSIS OF VARIANCE (ANOVA)

E.G

	A	B	C
	1	2	2
	2	4	3
	5	2	4
mean	2.67	2.67	3

1

$H_0 \rightarrow \mu_1 = \mu_2 = \mu_3$ → no diff in means

$H_a \rightarrow$ atleast 1 diff in the means

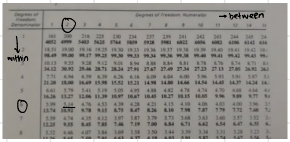
$\alpha = 0.5$ → generally

2

$df_{between} = K - 1 = 3 - 1 = 2$

$df_{within} = N - K = 9 - 3 = 6$

diff total $= 6 + 2 = 8$



$F_{critical} = 5.14$

3

$\bar{X}_1 = 2.67, \quad \bar{X}_2 = 2.67, \quad \bar{X}_3 = 3, \quad \bar{\bar{X}} = \frac{25}{9} = 2.78$

$$SS_{Total} = \Sigma (x - \bar{\bar{x}})^2$$
$$= (1-2.78)^2 + (2-2.78)^2 + (5-2.78)^2 + (2-2.78)^2 + (4-2.78)^2 + (2-2.78)^2 + (2-2.78)^2 + (3-2.78)^2 + (4-2.78)^2 = 13.6$$

$$SS_{within} = \Sigma (x_n - \bar{x}_n)^2$$
$$= (1-2.67)^2 + (2-2.67)^2 + (5-2.67)^2 + (2-2.67)^2 + (4-2.67)^2 + (2-2.67)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 = 13.34$$

$$SS_{between} = SS_{Total} - SS_{within}$$
$$= 13.6 - 13.34 = 0.23$$

4

$MS_{between} = \frac{SS_{between}}{df_{between}} = \frac{0.23}{2} = 0.12$

$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{13.34}{6} = 2.22$

1. State H_0, H_a, α
2. Degree's of freedom
   - ↳ df between groups $= K - 1$
   - ↳ df within groups $= N - K$ → no of groups; total no of observations
   - ↳ diff total $= N - 1$
   - ↳ F critical → using table b, w
3. Sum of square Deviations from mean
   - ↳ each groups mean → $\bar{x}_1, \bar{x}_2, \bar{x}_3$
   - ↳ Grand mean → $\bar{\bar{x}}$
   - ↳ $SS_{Total} = \Sigma (x - \bar{\bar{x}})^2$
   - ↳ $SS_{within} = \Sigma (x - \bar{x})^2$
   - ↳ $SS_{between} = SS_{Total} - SS_{within}$
4. Mean square
   - ↳ $MS_{between} = \frac{SS_{between}}{df_{between}}$
   - ↳ $MS_{within} = \frac{SS_{within}}{df_{within}}$
5. Calculate F statistic
   - ↳ $F = \frac{MS_{between}}{MS_{within}}$
   - ↳ conclusion

reject H_0	accept H_0
Fratio > Fcritical	Fratio < Fcritical
Fratio > 1	Fratio ≤ 1

5

$F = \frac{MS_{between}}{MS_{within}} = \frac{0.12}{2.22} = 0.05 \rightarrow < 1$

F critical $= 5.14$

$0.05 < 5.14$ and $F < 1$

hence accept H_0

CHP 13

13.1 Six different machines are being considered for use in manufacturing rubber seals. The machines are being compared with respect to tensile strength of the product. A random sample of four seals from each machine is used to determine whether the mean tensile strength varies from machine to machine. The following are the tensile-strength measurements in kilograms per square centimeter $\times 10^{-1}$:

Machine

1	2	3	4	5	6
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5
18.6	15.4	18.9	18.9	20.5	20.1

Perform the analysis of variance at the 0.05 level of significance and indicate whether or not the mean tensile strengths differ significantly for the six machines.

① $H_0 \rightarrow \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_1 \rightarrow$ atlesat 1 μ is diffent

$\alpha = 0.05$ $d_b = 5$ $P = 2.77$

$d_w = 18$

② $df_b = K-1 = 5$

$df_w = N-K = 24-6 = 18$

$df_t = 23$

③

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_T
17.2	17.175	18.175	17.75	18.425	18.025	17.79

④ $SS_T = 5.5003 + 9.6404 + 11.9004 + 21.6564 + 11.0804 + 8.2084$

$= 67.9863$

$SS_W = 4.1 + 8.1275 + 11.3075 + 21.65 + 9.4675 + 7.9875$

$= 62.64$

$S_b = 67.9863 - 62.64$

$= 5.3463$

$f = \dfrac{1.06926}{3.48} = 0.3072 < 1$ and < 2.77

accept Ho

⑤ $MS_b = \dfrac{5.3463}{5} = 1.06926$

$MS_w = \dfrac{62.64}{18} = 3.48$

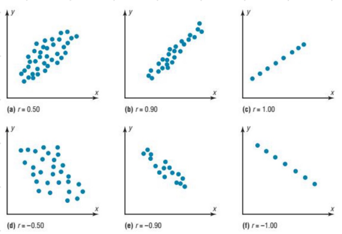
CORRELATION & REGRESSION

does a rs exist between variables

Scatter Plot

x - xais → independent variable → can be controlled

y- xais → dependent variable → cant be controlled



correlation coefficient

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

→ no of data pairs

Hypothesis testing for correlation

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$H_0 : \rho = 0$

$df = n-2$

correlation Coefficient (PPMC)

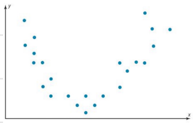
↳ if $r = 0$ then no correlation

↳ if $r \neq 0$ then correlation ($r > 0$, $x\uparrow \rightarrow y\uparrow$)

1. find r
2. state H_0, H_1
3. find t
4. find rejection region
5. If t lies in rejection region
   ↳ H_0 rejected
6. Type of rs

Type of rs

1. direct cause and effect rs (x causes y)
2. reverse cause and effect rs (y causes x)
3. caused by a third variable rs
4. curvilinear rs



Example # 05

- Test the significance of the correlation coefficient found in Example # 01.

Subject	Age, x	Pressure, y	xy	x^2	y^2
A	43	128	5,504	1,849	16,384
B	48	120	5,760	2,304	14,400
C	56	135	7,560	3,136	18,225
D	61	143	8,723	3,721	20,449
E	67	141	9,447	4,489	19,881
F	70	152	10,640	4,900	23,104
	$\Sigma x = 345$	$\Sigma y = 819$	$\Sigma xy = 47{,}634$	$\Sigma x^2 = 20{,}399$	$\Sigma y^2 = 112{,}443$

$H_0: \rho = 0 \qquad H_1: \rho \neq 0$

$$r = \frac{6(47{,}634) - (345)(819)}{\sqrt{[6(20399) - (345)^2]\,[6(112{,}443) - (819)^2]}} = 0.897$$

$$t = 0.897\sqrt{\frac{6-2}{1-0.897^2}} = 4.059$$

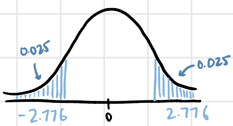
why

df: $6-2 = 4 \qquad \alpha: 0.025$

t lies inside rejection region

H_0 rejected

so there is correlation



df \ α	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	318.3
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	22.33
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	10.21
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.787

Stopping Distances

In a study on speed control, it was found that the main reasons for regulations were to make traffic flow more efficient and to minimize the risk of danger. An area that was focused on in the study was the distance required to completely stop a vehicle at various speeds. Use the following table to answer the questions.

MPH	Braking distance (feet)
20	20
30	45
40	81
50	133
60	205
80	411

Assume MPH is going to be used to predict stopping distance.

1. Which of the two variables is the independent variable?
2. Which is the dependent variable?
3. What type of variable is the independent variable?
4. What type of variable is the dependent variable?
5. Construct a scatter plot for the data.
6. Is there a linear relationship between the two variables?
7. Is the relationship positive or negative?
8. Can braking distance be accurately predicted from MPH?
9. List some other variables that affect braking distance.
10. Compute the value of r.
11. Is r significant at $\alpha = 0.05$?

Solve How

Simple Linear Regression

$y = mx + c$

↓

regression

Example # 06

• Find the equation of the regression line for the data in Example # 01 (Slide 15), and graph the line on the scatter plot of the data.

Subject	Age, x	Pressure, y	xy	x^2	y^2
A	43	128	5,504	1,849	16,384
B	48	120	5,760	2,304	14,400
C	56	135	7,560	3,136	18,225
D	61	143	8,723	3,721	20,449
E	67	141	9,447	4,489	19,881
F	70	152	10,640	4,900	23,104
	$\Sigma x = 345$	$\Sigma y = 819$	$\Sigma xy = 47{,}634$	$\Sigma x^2 = 20{,}399$	$\Sigma y^2 = 112{,}443$

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{(819)(20{,}399) - (345)(47{,}634)}{(6)(20{,}399) - (345)^2} = 81.048$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{(6)(47{,}634) - (345)(819)}{(6)(20{,}399) - (345)^2} = 0.964$$

$$y' = 81.048 + 0.964x$$

sign of the correlation coefficient and the sign of the slope of the regression line

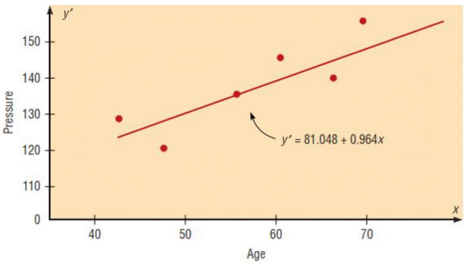
Prediction Using Regression Eq.

• Using the equation of the regression line found in Example # 06, predict the blood pressure for a person who is 50 years old.

Substituting 50 for x in the regression line $y' = 81.048 + 0.964x$ gives

$$y' = 81.048 + (0.964)(50) = 129.248 \text{ (rounded to 129)}$$

In other words, the predicted systolic blood pressure for a 50-year-old person is 129.



$y = mx + c$

Method of least sq

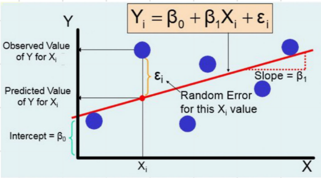
$$B_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$B_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

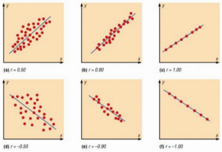
random error

$$Y = \beta_0 + \beta_1 x + \epsilon$$

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$



Relationship B/w Regression & Correlation



The value of the correlation coefficient can also be found by using the formula:

$$r = \frac{bs_x}{s_y}$$

CHP 11

11.3 The amounts of a chemical compound y that dissolved in 100 grams of water at various temperatures x were recorded as follows:

x (°C)	y (grams)		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	42
75	48	51	44

(a) Find the equation of the regression line.
(b) Graph the line on a scatter diagram.
(c) Estimate the amount of chemical that will dissolve in 100 grams of water at 50°C.

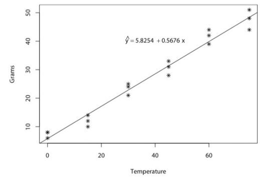
11.3 (a) $\sum_i x_i = 675$, $\sum_i y_i = 488$, $\sum_i x_i^2 = 37,125$, $\sum_i x_i y_i = 25,005$, $n = 18$. Therefore,

HOW

$$b = \frac{(18)(25,005) - (675)(488)}{(18)(37,125) - (675)^2} = 0.5676,$$
$$a = \frac{488 - (0.5676)(675)}{18} = 5.8254.$$

Hence $\hat{y} = 5.8254 + 0.5676x$

(b) The scatter plot and the regression line are shown below.



(c) For $x = 50$, $\hat{y} = 5.8254 + (0.5676)(50) = 34.205$ grams.

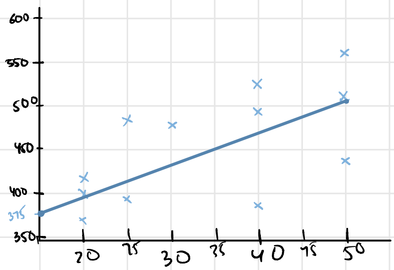
$\Sigma x = 225$ $\quad \Sigma x^2 = 12375$ $\quad \Sigma xy =$

$\Sigma y = 488$ $\quad \Sigma y^2$

11.9 A study was made by a retail merchant to determine the relation between weekly advertising expenditures and sales.

Advertising Costs ($)	Sales ($)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

(a) Plot a scatter diagram.
(b) Find the equation of the regression line to predict weekly sales from advertising expenditures.
(c) Estimate the weekly sales when advertising costs are $35.
(d) Plot the residuals versus advertising costs. Comment.



b) $\Sigma x = 410$ $\quad \Sigma x^2 = 15,650$ $\quad \Sigma xy = 191325$

$\Sigma y = 5445$ $\quad n = 12$

$$B_0 = \frac{(5445)(15650) - (410)(191325)}{12(15650) - 410^2} = 343.7056$$

$$B_1 = \frac{12(191325) - (410)(5445)}{12(15650) - 410^2} = 3.22081$$

$y = 343.7056 + 3.22081x$

c) $y = 343.7056 + 3.22081(35)$

$y = 456.434$

CHP 11

11.43 Compute and interpret the ~~correlation coefficient for the following grades of 6 students selected at~~ random:

Mathematics grade	70	92	80	74	65	83
English grade	74	84	63	87	78	90

x	x^2	y	y^2	xy
464	36354	476	38254	36926

$$r = \frac{6(36926)-(464)(476)}{\sqrt{[6(36354)-(464)^2][6(38254)-(476)^2]}}$$

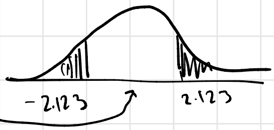
$r = 0.2396$

11.46 Test the hypothesis that $\rho = 0$ in Exercise 11.43 against the alternative that $\rho \neq 0$. Use a 0.05 level of significance.

$H_0 : P = 0$ $\quad \alpha = 0.05$ $\quad n = 6$

$H_1 : P \neq 0$ $\quad r = 0.2396$

$$t_c = 0.2396\sqrt{\frac{6-2}{1-0.2396^2}} = 0.4935$$



df = 6-2 = 4, t from table = 2.123 → rejection region

DONT REJECT Ho

11.45 With reference to Exercise 11.13 on page 400, assume a bivariate normal distribution for x and y.

(a) Calculate r.

(b) Test the null hypothesis that $\rho = -0.5$ against the alternative that $\rho < -0.5$ at the 0.025 level of significance.

(c) Determine the percentage of the variation in the amount of particulate removed that is due to changes in the daily amount of rainfall.

Daily Rainfall, x (0.01 cm)	Particulate Removed, y ($\mu g/m^3$)
4.3	126
4.5	121
5.9	116
5.6	118
6.1	114
5.2	118
3.8	132
2.1	141
7.5	108

a) .... $r = -0.919$

$$t = -0.919\sqrt{\frac{9-2}{1-(0.919)^2}} = -12.71$$

WHY z test

marginal.

11.45 (a) From the data of Exercise 11.9 we find $S_{xx} = 244.26 - 45^2/9 = 19.26$, $S_{yy} = 133,786 - 1094^2/9 = 804.2222$, and $S_{xy} = 5348.2-(45)(1094)/9 = -121.8$. So, $r = \frac{-121.8}{\sqrt{(19.26)(804.2222)}} = -0.979$.

(b) The hypotheses are

$$H_0 : \rho = -0.5,$$
$$H_1 : \rho < -0.5.$$

$\alpha = 0.025$.
Critical regions: $z < -1.96$.
Computations: $z = \frac{\sqrt{6}}{2} \ln\left[\frac{(0.021)(1.5)}{(1.979)(0.5)}\right] = -4.22$.
Decision: Reject H_0: $\rho < -0.5$.

(c) $(-0.979)^2(100\%) = 95.8\%$.

ishma hafeez notes

BAYES RULE

$$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

↳ $P(A|B) = \frac{P(A \cap B)}{P(B)}$ $\qquad$ $P(B|A) = \frac{P(A\cap B)}{P(A)}$

↳ $P(A|B)\,P(B) = P(B|A)\,P(A)$

Joint Probability Distribution

1. $f(x,y) \geq 0$
2. $\sum_x \sum_y f(x,y) = 1$
3. $P(X=x, Y=y) = f(x,y)$

↳ $P[(X,Y) \in A] = \sum\sum_A f(x,y)$

$\sigma^2 \rightarrow$ variance

$V(X) = E(X^2) - [E(X)]^2$

$V(Y) = E(Y^2) - [E(Y)]^2$

mean

$\mu = E(X) = \sum_x x f(x)$

$E(XY) = \sum\sum xy\ f(x,y)$

Covariance $(X,Y) = E(XY) - E(X).E(Y)$

Corelation $(X,Y) = \frac{\text{Covariance }(X,Y)}{\sqrt{V(X)}\ \sqrt{V(Y)}}$

$\pm 1 \longrightarrow$ strong
$\pm 0.5 \longrightarrow$ moderate
$0 \longrightarrow$ negligible

$IQR = Q_3 - Q_1$

$L = Q_1 - 1.5(IQR)$

$U = Q_3 + 1.5(IQR)$

Probabilily mass Function (PMF)

1. $f(x) \geq 0$, for all $x \in R$

↳ $\sum f(x)\,dx = 1$ → verify PMF

↳ $P(X=x) = F(x) - F(x-1)$

↳ $P(a < x < b) = F(b) - F(a)$

$P(n) = (n+1)^n$

Emperical Rule

68%	$\bar{x} \pm s$
95%	$\bar{x} \pm 2s$
99.7%	$\bar{x} \pm 3s$

Standard Normal Distribution

$$Z\text{-score} = \frac{x - \mu}{\sigma}$$

Centeral limit theorem

z test

$$Z = \frac{\bar{x} - \mu}{\underbrace{\sigma\sqrt{n}}_{\text{standard error of mean}}}$$

1. $\bar{x} = \frac{\sum x}{n}$ — $\bar{x}$: sample mean; n: sample size
2. $S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$ — S: S.D
3. $S_{\bar{x}} = \frac{s}{\sqrt{n}}$ — $S_{\bar{x}}$: standard error mean; s: S.D
4. $E = \underbrace{Z_{\frac{\alpha}{2}}}_{\text{critical region of z-distribution}} \underbrace{\left(\frac{\sigma}{\sqrt{n}}\right)}_{\text{standard error of mean}}$ — E: margin of error
5. $n = \left(\frac{Z_{\frac{\alpha}{2}}\,\sigma}{E}\right)^2$

Finite Population Correction Factor

$$\sqrt{\frac{N - n}{N - 1}}$$

N: Population size; n: sample size

Confidence Inteval for Propotions

1. $\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$\hat{p}$: sample Proportion; $\hat{p} = \frac{x}{n}$, $x \rightarrow$ no of sample units; $\hat{q} \rightarrow 1 - \hat{p}$

Chi Square Distribution

1. $\frac{(n-1)s^2}{\chi^2_{right}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{left}}$
2. $df = n - 1$

Confidence Interval for μ

$\rightarrow$ 1 distribution

2. $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ — s: sample S.D
   - ↳ σ unknown
   - ↳ $n \geq 30$
3. $\bar{x} \pm t_{\frac{\alpha}{2}(v)} \frac{s}{\sqrt{n}}$ — s: sample S.D; $t_{\frac{\alpha}{2}(v)}$: critical value of t distribution; v: n-1
   - ↳ σ unknown
   - ↳ $n < 30$

Confidence Interval for $\mu_1 - \mu_2$

$\rightarrow$ 2 distributions

2. $(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
   - ↳ σ_1, σ_2 unknown
   - ↳ $n \geq 30$
3. $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
   - ↳ σ_1, σ_2 unknown
   - ↳ $n < 30$

Z-Test

↳ $n \geq 30$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z = \frac{(\underbrace{\bar{x}_1 - \bar{x}_2}_{\text{sample mean}}) - (\underbrace{\mu_1 - \mu_2}_{H_0\text{ mean}})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standard deviation: σ_1^2, σ_2^2 — sample size: n_1, n_2

$$\delta = \sqrt{\frac{n \Sigma x^2 - (\Sigma x)^2}{n(n-1)}}$$

T-Test for mean

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$df = n-1$

↳ $\sigma_1 \neq \sigma_2$ ← variance

$$t = \frac{(\underbrace{\bar{x}_1 - \bar{x}_2}_{\text{sample mean}}) - (\underbrace{\mu_1 - \mu_2}_{H_0\text{ mean}})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standard deviation: σ_1^2, σ_2^2 — sample size: n_1, n_2

$$df = \frac{(\delta_1 + \delta_2)^2}{\frac{\delta_1^2}{n-1} + \frac{\delta_2^2}{n-1}}$$

↳ $\sigma_1 = \sigma_2$ ← variance

$$\delta = \frac{(n-1)(\delta_1)^2 + (n-1)(\delta_2)^2}{(n+n)-2}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\delta \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\delta = \sqrt{\frac{n \Sigma x^2 - (\Sigma x)^2}{n(n-1)}}$$

1. State Ho, Ha, α
2. Degree's of freedom
   - ↳ df between groups $= K-1$
   - ↳ df within groups $= N - K$ ← no. of groups; ↳ total no of observations
   - ↳ diff total $= N-1$
   - ↳ F critical → using table
3. Sum of square deviations from mean
   - ↳ each groups mean → $\bar{x}_1, \bar{x}_2, \bar{x}_3$
   - ↳ Grand mean → $\bar{\bar{x}}$
   - ↳ $SS_{Total} = \Sigma(x - \bar{\bar{x}})^2$
   - ↳ $SS_{within} = \Sigma(x - \bar{x})^2$
   - ↳ $SS_{between} = SS_{Total} - SS_{within}$
4. Mean square
   - ↳ $MS_{between} = \frac{SS_{between}}{df_{between}}$
   - ↳ $MS_{within} = \frac{SS_{within}}{df_{within}}$
5. Calculate F statistic
   - ↳ $F = \frac{MS_{between}}{MS_{within}}$
   - ↳ conclusion

reject Ho	accept Ho
Fratio > Fcritical	Fratio < Fcritical
Fratio > 1	Fratio ≤ 1

correlation coefficient (PPMC)

correlation coefficient

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

↳ no. of data pairs

Hypothesis testing for correlation

$$t = r\sqrt{\frac{n-2}{1-r^2}} \qquad H_0: \rho = 0$$

$df = n-2$

simple linear regression

method of least sq

$$\beta_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} \qquad \beta_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$Y = \beta_0 + \beta_1 x + \underbrace{\epsilon}_{\text{random error}}$$