Comparative Experiments

Example 2

Who is a better ODI batsman, Virat or Babar? (Based on the runs sc

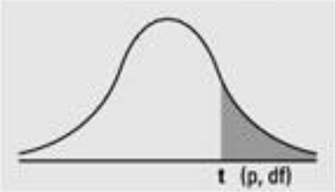
Batsman	One sample each of 10 ODI innings
Virat	00, 53, 34, 31, 00, 54, 96, 20, 10, 19
Babar	12, 09, 91, 79, 51, 45, 41, 46, 29, 33

What is the hypothesis test?

What is the statistical (mathematical) model based on the hy

What's the statistical conclusion?

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



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	0.270722	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	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	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	0.257599	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

Example 3

Given that 9 bearings made by a certain process have an average diameter of 0.305 cm and the sample standard deviation of 0.003 cm, construct a 99 % confidence interval for the true mean diameter of bearings made by the process. What is the width of the confidence interval?

Degrees of freedom (ν)	Amount of area in one tail (α)							
	0.0005	0.001	0.005	0.010	0.025	0.050	0.100	0.200
1	636.6192	318.3088	63.65674	31.82052	12.70620	6.313752	3.077684	1.376382
2	31.59905	22.32712	9.924843	6.964557	4.302653	2.919986	1.885618	1.060660
3	12.92398	10.21453	5.840909	4.540703	3.182446	2.353363	1.637744	0.978472
4	8.610302	7.173182	4.604095	3.746947	2.776445	2.131847	1.533206	0.940965
5	6.868827	5.893430	4.032143	3.364930	2.570582	2.015048	1.475884	0.919544
6	5.958816	5.207626	3.707428	3.142668	2.446912	1.943180	1.439756	0.905703
7	5.407883	4.785290	3.499483	2.997952	2.364624	1.894579	1.414924	0.896030
8	5.041305	4.500791	3.355387	2.896459	2.306004	1.859548	1.396815	0.888890
9	4.780913	4.296806	3.249836	2.821438	2.262157	1.833113	1.383029	0.883404
10	4.586894	4.143700	3.169273	2.763769	2.228139	1.812461	1.372184	0.879058
11	4.436979	4.024701	3.105807	2.718079	2.200985	1.795885	1.363430	0.875530
12	4.317791	3.929633	3.054540	2.680998	2.178813	1.782288	1.356217	0.872609
13	4.220832	3.851982	3.012276	2.650309	2.160369	1.770933	1.350171	0.870152
14	4.140454	3.787390	2.976843	2.624494	2.144787	1.761310	1.345030	0.868055
15	4.072765	3.732834	2.946713	2.602480	2.131450	1.753050	1.340606	0.866245
16	4.014996	3.686155	2.920782	2.583487	2.119905	1.745884	1.336757	0.864667
17	3.965126	3.645767	2.898231	2.566934	2.109816	1.739607	1.333379	0.863279

https://www.mathsisfun.com/data/standard-normal-distribution-table.html

Choice of Sample Size

- Selection of appropriate sample size 'n' is critical in any experimental design
- In the previous example, have a look at the length of 100*(1-α)% confidence interval for difference in means $(\mu_1 - \mu_2)$
- It was determined by

$$t_{\alpha/2,\, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

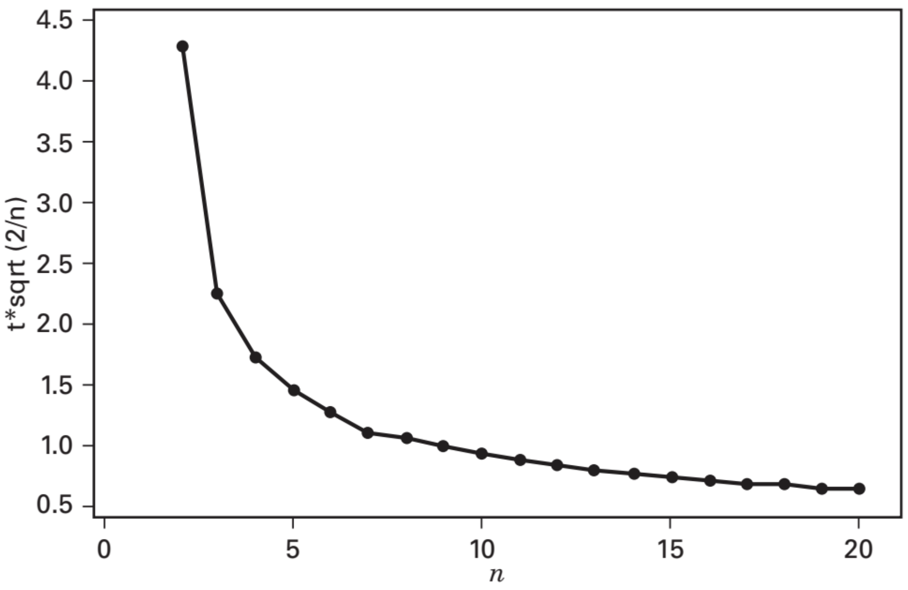
$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2$$

$$\leq \bar{y}_1 - \bar{y}_2 + t_{\alpha/2, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

is a $100(1-\alpha)$ percent confidence interval for $\mu_1 - \mu_2$.

- What is the effect of sample size on this width?
  - Say n1 = n2 = n, and α = 0.05, Sp could be anything (we don't have control over it)
  - So essentially, the width is a function of

$$t_{\alpha/2,\, 2n-2}\, S_p \sqrt{\frac{2}{n}}$$



Assumptions in t-test

- In using the t-test procedure, we make the assumption that
  - both samples are *random samples that are drawn from independent populations with normal distribution*, and
  - *the standard deviation or variances of both populations are equal.*
- **The assumption of independence is critical**, and if the run order is randomized (and, if appropriate, other experimental units and materials are selected at random), this assumption will usually be satisfied.
- The equal variance and normality assumptions are easy to check using **a normal probability plot.**

Normal Probability Plot

- The equal variance and normality assumptions are easy to check using **a normal probability plot.**

To construct the Normal Probability Plot

- First the sample $y_1, y_2, y_3, \dots, y_n$ is arranged in the increasing order $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ where $y_{(1)}$ is the smallest observation and $y_{(n)}$ is the largest observation
- These ordered observations $y_{(i)}$ are plotted on X-axis
- On the Y-axis, we plot their cumulative frequency (i-0.5)/n

(empirically, it should be = i/n, but we use correction for discrete data)

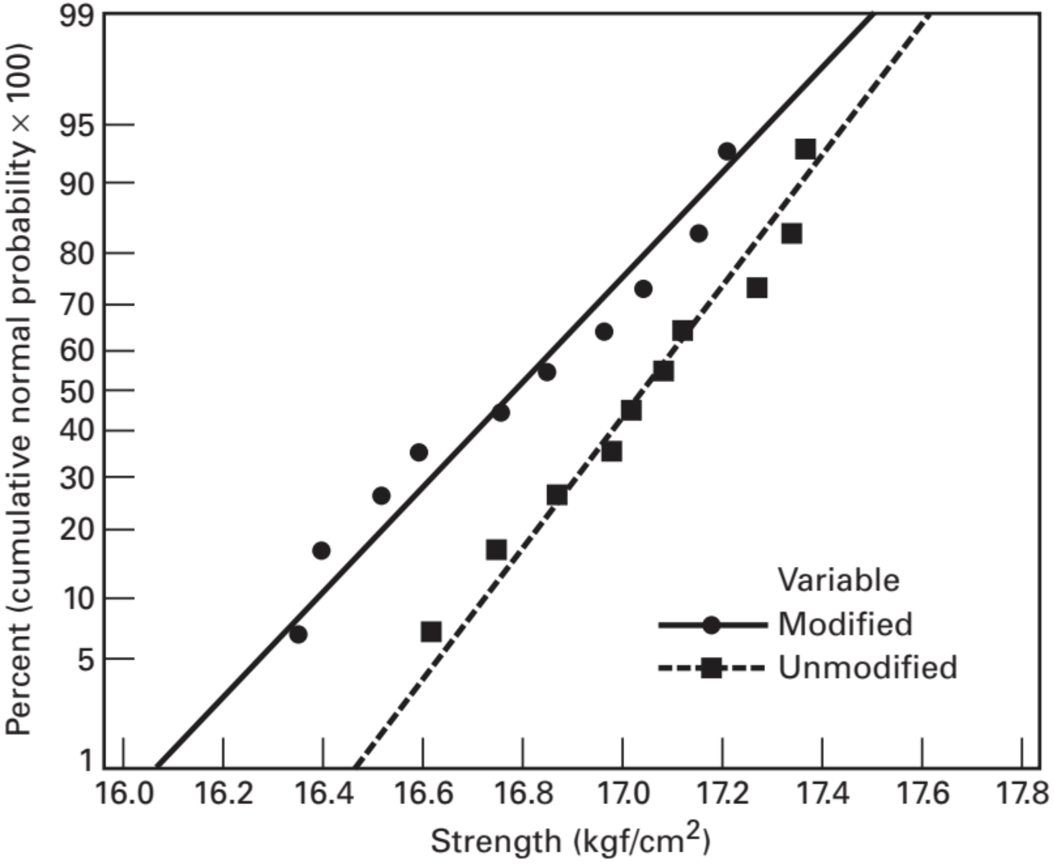
- Then you arrange the Y-axis so that if the hypothesized distribution

adequately describes the data, the plotted points will follow a

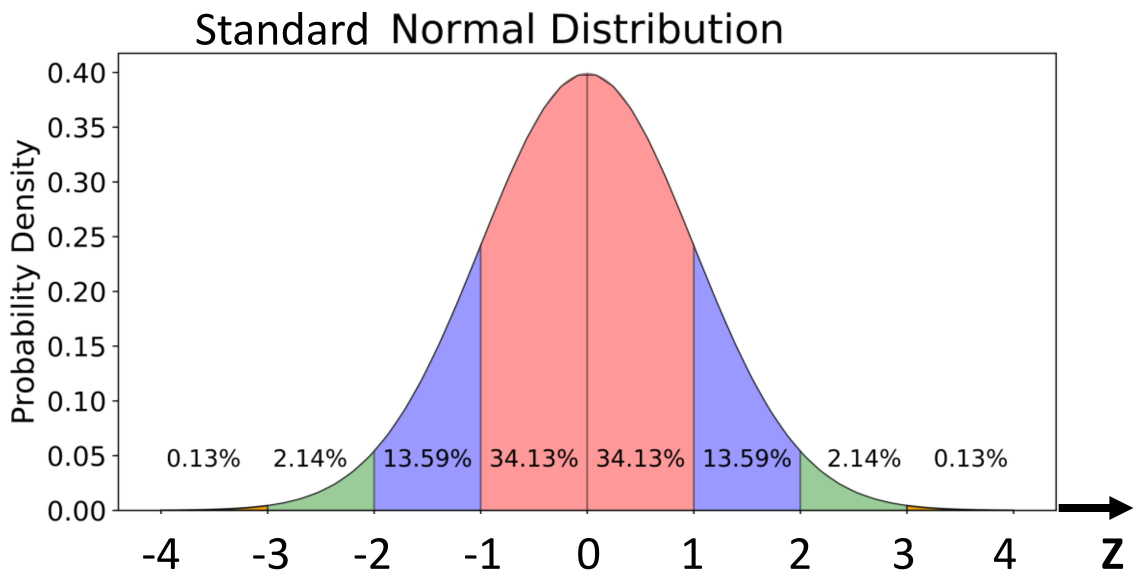
Straight line

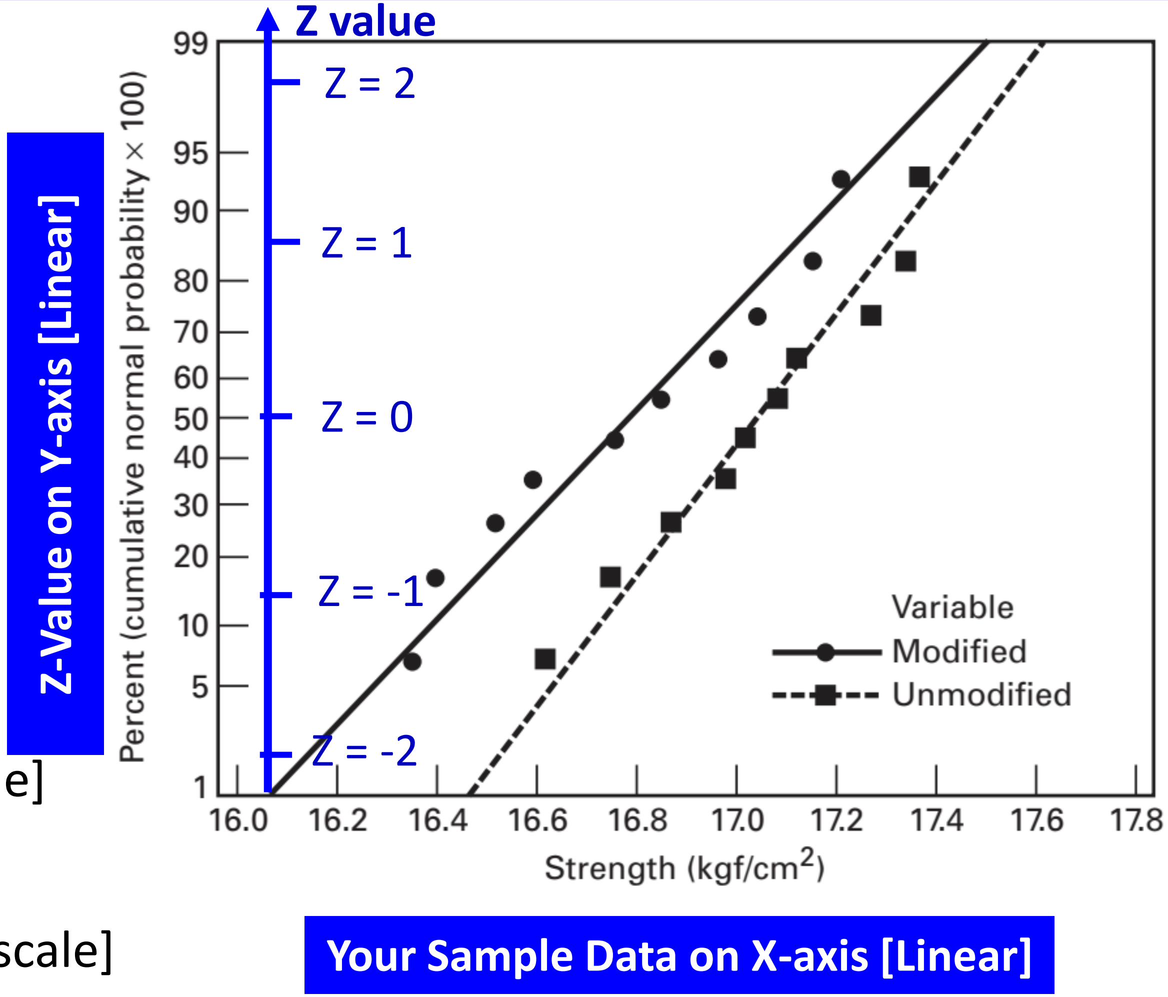
- If the slopes of both the lines is approx. same, then the assumption

of equal variances is valid



Recap: How to Plot Normal Probability Plot





Normal Probability Plot Construction

- On X-axis: Sample data [Linear scale]
- On Y-axis: Find Z-value for a particular data point

Z-value = Z(CDF of X_i) = Z ((i-0.5)/n) [Linear scale]

(Note, if you show CDF values on Y-axis, the scale is non-linear)

Example 4

CEP2022_Notebook (2.1.6)

Nerve preservation is important in surgery because accidental injury to the nerve can lead to post-surgical problems such as numbness, pain, or paralysis. Nerves are usually identified by their appearance and relationship to nearby structures or detected by local electrical stimulation (electromyography), but it is relatively easy to overlook them.

An article in Nature Biotechnology ("Fluorescent Peptides Highlight Peripheral Nerves During Surgery in Mice," Vol. 29, 2011) describes the use of a fluorescently labeled peptide that binds to nerves to assist in identification. Table 2.3 shows the normalized fluorescence after two hours for nerve and muscle tissue for 12 mice (the data were read from a graph in the paper).

TABLE 2.3
Normalized Fluorescence After Two Hours

Observation	Nerve	Muscle
1	6625	3900
2	6000	3500
3	5450	3450
4	5200	3200
5	5175	2980
6	4900	2800
7	4750	2500
8	4500	2400
9	3985	2200
10	900	1200
11	450	1150
12	2800	1130

Example 4

- Assuming a common variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$ **(??)**

Hypothesis Testing

$$H_0: \mu_1 \leq \mu_2$$
$$H_1: \mu_1 > \mu_2$$

TABLE 2.3
Normalized Fluorescence After Two Hours

Observation	Nerve	Muscle
1	6625	3900
2	6000	3500
3	5450	3450
4	5200	3200
5	5175	2980
6	4900	2800
7	4750	2500
8	4500	2400
9	3985	2200
10	900	1200
11	450	1150
12	2800	1130

DIY

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	0.270722	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	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	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	0.257599	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	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	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
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	———	———	80%	90%	95%	98%	99%	99.9%

Is our Assumption Correct?

Is it okay to assume common variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$?

Normal Probability Plot