

MATH 450: Mathematical statistics

November 17th, 2020

Lecture 21: P-values

Week 2	•	Chapter 6: Statistics and Sampling Distributions
Week 4	•	Chapter 7: Point Estimation
Week 7	•	Chapter 8: Confidence Intervals
Week 10	•	Chapter 9: Tests of Hypotheses
Week 12	•	Chapter 10: Two-sample testing
Week 14	•	Regression

Key steps in statistical inference

- Understand statistical models [Chapter 6]
- Come up with reasonable estimates of the parameters of interest [Chapter 7]
- Quantify the confidence with the estimates [Chapter 8]
- Testing with the parameters of interest [Chapter 9]

Contexts

- The central mega-example: population mean μ
- Difference between two population means

9.1 Hypotheses and test procedures

- test procedures
- errors in hypothesis testing
- significance level

9.2 Tests about a population mean

- normal population with known σ
- large-sample tests
- a normal population with unknown σ

9.4 P-values

Hypothesis testing for one parameter

- 1 Identify the parameter of interest
- 2 Determine the null value and state the null hypothesis
- 3 State the appropriate alternative hypothesis
- 4 Give the formula for the test statistic
- 5 State the rejection region for the selected significance level α
- 6 Compute statistic value from data
- 7 Decide whether H_0 should be rejected and state this conclusion in the problem context

Test about a population mean

- Null hypothesis

$$H_0 : \mu = \mu_0$$

- The alternative hypothesis will be either:
 - $H_a : \mu > \mu_0$
 - $H_a : \mu < \mu_0$
 - $H_a : \mu \neq \mu_0$

Normal population with known σ

Null hypothesis: $\mu = \mu_0$

Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

...

Alternative Hypothesis

$$H_a: \mu > \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu \neq \mu_0$$

Rejection Region for Level α Test

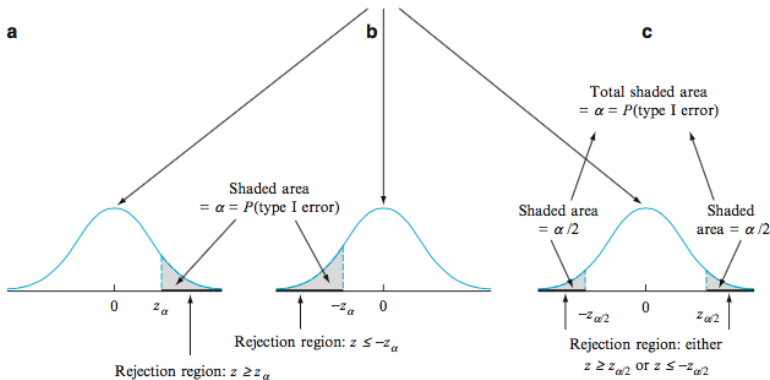
$$z \geq z_\alpha \text{ (upper-tailed test)}$$

$$z \leq -z_\alpha \text{ (lower-tailed test)}$$

$$\text{either } z \geq z_{\alpha/2} \text{ or } z \leq -z_{\alpha/2} \text{ (two-tailed test)}$$

General rule

z curve (probability distribution of test statistic Z when H_0 is true)



Large-sample tests

Null hypothesis: $\mu = \mu_0$

Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

...

Alternative Hypothesis

$$H_a: \mu > \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu \neq \mu_0$$

Rejection Region for Level α Test

$$z \geq z_\alpha \text{ (upper-tailed test)}$$

$$z \leq -z_\alpha \text{ (lower-tailed test)}$$

$$\text{either } z \geq z_{\alpha/2} \text{ or } z \leq -z_{\alpha/2} \text{ (two-tailed test)}$$

[Does not need the normal assumption]

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic value: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Alternative Hypothesis

$H_a: \mu > \mu_0$

$H_a: \mu < \mu_0$

$H_a: \mu \neq \mu_0$

Rejection Region for a Level α Test

$t \geq t_{\alpha, n-1}$ (upper-tailed)

$t \leq -t_{\alpha, n-1}$ (lower-tailed)

either $t \geq t_{\alpha/2, n-1}$ or $t \leq -t_{\alpha/2, n-1}$ (two-tailed)

[Require normal assumption]

P-values

- The common approach in statistical testing is:
 - 1 specifying significance level α
 - 2 reject/not reject H_0 based on evidence
- Weaknesses of this approach:
 - it says nothing about whether the computed value of the test statistic just barely fell into the rejection region or whether it exceeded the critical value by a large amount
 - each individual may select their own significance level for their presentation
- We also want to include some *objective* quantity that describes how *strong* the rejection is \rightarrow P-value

Problem

Suppose μ was the true average nicotine content of brand of cigarettes. We want to test:

$$H_0 : \mu = 1.5$$

$$H_a : \mu > 1.5$$

Suppose that $n = 64$ and $z = \frac{\bar{x} - 1.5}{s/\sqrt{n}} = 2.1$. Will we reject H_0 if the significance level is

- (a) $\alpha = 0.05$
- (b) $\alpha = 0.025$
- (c) $\alpha = 0.01$
- (d) $\alpha = 0.005$

Level of Significance α	Rejection Region	Conclusion
.05	$z \geq 1.645$	Reject H_0
.025	$z \geq 1.96$	Reject H_0
.01	$z \geq 2.33$	Do not reject H_0
.005	$z \geq 2.58$	Do not reject H_0

Question: What is the smallest value of α for which H_0 is rejected.

DEFINITION

The ***P*-value** (or *observed significance level*) is the smallest level of significance at which H_0 would be rejected when a specified test procedure is used on a given data set. Once the *P*-value has been determined, the conclusion at any particular level α results from comparing the *P*-value to α :

1. $P\text{-value} \leq \alpha \Rightarrow$ reject H_0 at level α .
2. $P\text{-value} > \alpha \Rightarrow$ do not reject H_0 at level α .

Testing by P-value method

DECISION
RULE BASED
ON THE
P-VALUE

Select a significance level α (as before, the desired type I error probability).
Then reject H_0 if $P\text{-value} \leq \alpha$; do not reject H_0 if $P\text{-value} > \alpha$

Remark: the smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis.

P-values for z-tests

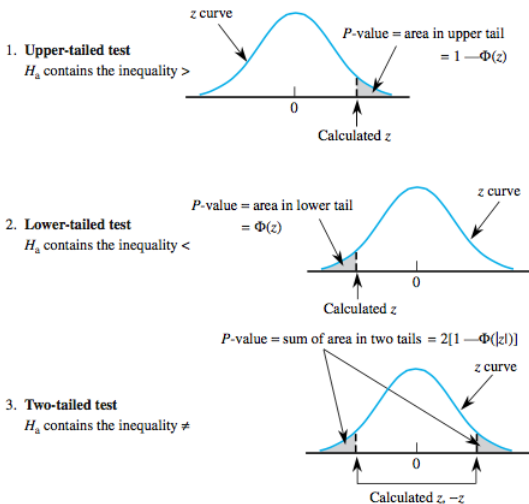


Figure 9.7 Determination of the P -value for a z test

Problem

The target thickness for silicon wafers used in a certain type of integrated circuit is $245 \mu\text{m}$. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of $246.18 \mu\text{m}$ and a sample standard deviation of $3.60 \mu\text{m}$.

At confidence level $\alpha = 0.01$, does this data suggest that true average wafer thickness is something other than the target value?

$\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.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

P-values for z-tests

1. Parameter of interest: $\mu =$ true average wafer thickness
2. Null hypothesis: $H_0: \mu = 245$
3. Alternative hypothesis: $H_a: \mu \neq 245$
4. Formula for test statistic value: $z = \frac{\bar{x} - 245}{s/\sqrt{n}}$
5. Calculation of test statistic value: $z = \frac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$
6. Determination of P -value: Because the test is two-tailed,
$$P\text{-value} = 2[1 - \Phi(2.32)] = .0204$$
7. Conclusion: Using a significance level of .01, H_0 would not be rejected since $.0204 > .01$. At this significance level, there is insufficient evidence to conclude that true average thickness differs from the target value.

P-values for z-tests

$$P\text{-value: } P = \begin{cases} 1 - \Phi(z) & \text{for an upper-tailed test} \\ \Phi(z) & \text{for a lower-tailed test} \\ 2[1 - \Phi(|z|)] & \text{for a two-tailed test} \end{cases}$$

P-values for t -tests

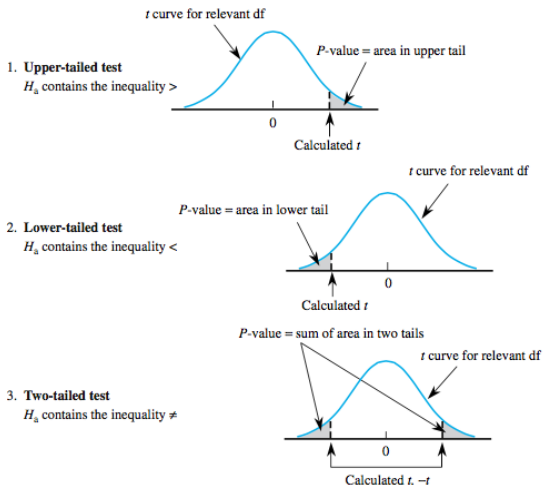


Figure 9.8 P -values for t tests

Problem

Suppose we want to test

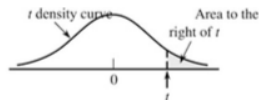
$$H_0 : \mu = 25$$

$$H_a : \mu > 25$$

from a sample with $n = 5$ and the calculated value

$$t = \frac{\bar{x} - 25}{s/\sqrt{n}} = 1.02$$

- (a) *What is the P-value of the test*
- (b) *Should we reject the null hypothesis?*

Table A.7 t Curve Tail Areas

t	ν	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0.0		.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500
0.1		.468	.465	.463	.463	.462	.462	.462	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461
0.2		.437	.430	.427	.426	.425	.424	.424	.423	.423	.423	.422	.422	.422	.422	.422	.422	.422	.422
0.3		.407	.396	.392	.390	.388	.387	.386	.386	.386	.385	.385	.385	.384	.384	.384	.384	.384	.384
0.4		.379	.364	.358	.355	.353	.352	.351	.350	.349	.349	.348	.348	.348	.347	.347	.347	.347	.347
0.5		.352	.333	.326	.322	.319	.317	.316	.315	.315	.314	.313	.313	.313	.312	.312	.312	.312	.312
0.6		.328	.305	.295	.290	.287	.285	.284	.283	.282	.281	.280	.280	.279	.279	.279	.278	.278	.278
0.7		.306	.278	.267	.261	.258	.255	.253	.252	.251	.250	.249	.249	.248	.247	.247	.247	.247	.246
0.8		.285	.254	.241	.234	.230	.227	.225	.223	.222	.221	.220	.220	.219	.218	.218	.218	.217	.217
0.9		.267	.232	.217	.210	.205	.201	.199	.197	.196	.195	.194	.193	.192	.191	.191	.191	.190	.190
1.0		.250	.211	.196	.187	.182	.178	.175	.173	.172	.170	.169	.169	.168	.167	.167	.166	.166	.165
1.1		.235	.193	.176	.167	.162	.157	.154	.152	.150	.149	.147	.146	.146	.144	.144	.144	.143	.143
1.2		.221	.177	.158	.148	.142	.138	.135	.132	.130	.129	.128	.127	.126	.124	.124	.124	.123	.123
1.3		.209	.162	.142	.132	.125	.121	.117	.115	.113	.111	.110	.109	.108	.107	.107	.106	.105	.105
1.4		.197	.148	.128	.117	.110	.106	.102	.100	.098	.096	.095	.093	.092	.091	.091	.090	.090	.089
1.5		.187	.136	.115	.104	.097	.092	.089	.086	.084	.082	.081	.080	.079	.077	.077	.077	.076	.075

A P-value:

- is not the probability that H_0 is true
- is not the probability of rejecting H_0
- is the probability, calculated assuming that H_0 is true, of obtaining a test statistic value at least as contradictory to the null hypothesis as the value that actually resulted