

Inference with Means

(one sample)

AP Statistics
Chapter 23

William Gossett

1876 - 1937

Guinness employee



Dublin, Ireland

“Quality Assurance”
(a.k.a., taste tester)
for **Guinness beer**
(HORRIBLE JOB, RIGHT?)

William Gossett

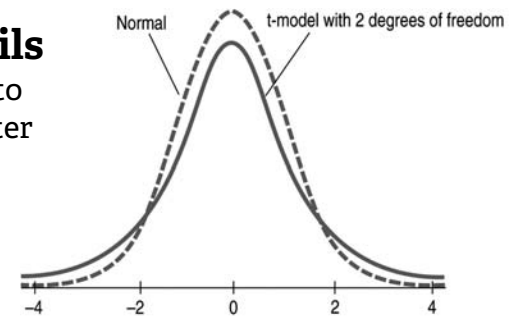


- Taste-tested batches of dark lager in samples
- Calculated that he should reject good beer about 5% of the time
- Actually rejected good beer about 15% of the time **(WHAT?!!)**
- Found that the Normal model doesn't play nice with small samples...

What Does This **mean** for **Means**?

The Student's *t*-models

These curves have **more area in the tails**
(Because we are using “s” to estimate “ σ ”, there is greater “uncertainty”/variability)



What Does This **mean** for **Means**?

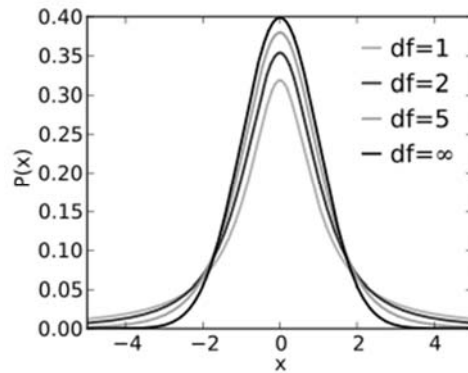
The Student's *t*-models

Degrees of freedom (*df*)

Each sample size has its own model/curve!

For 1-sample means:

$$df = n - 1$$



Student's *t*-models

(you want to write this stuff down...)

- Use when we don't know the population standard deviation σ (when we use the sample standard deviation as an estimate for σ)
- *t*-distribution has fatter tails than *z*-distribution
- As *df* increase, the *t*-models look more and more like the Normal model.
- In fact, the *t*-model with infinite degrees of freedom is exactly Normal.

Calculating *t* vs *z*

(you want to write this down too...)

If we know σ :
(population SD)

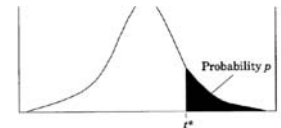
If we only know *s*:
(sample SD)

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

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

t-table Practice

Table entry for *p* and *C* is the point * with probability *p* lying above it and probability *C* lying between $-t^*$ and t^* .



Find the critical value of *t* for 95% confidence with...

a) *df* = 10

$$t^* = 2.228$$

b) *n* = 20

$$t^* = 2.093$$

(use *df* = 20 - 1 = 19)

c) *df* = 32

$$t^* = 2.042$$

(use *df* = 30 - round DOWN)

Table B

df	Tail probability <i>p</i>									
	.25	.20	.15	.10	.05	.025	.01	.005	.001	.0005
1	1.000	1.375	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.33
2	.806	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598
5	.727	.929	1.156	1.476	2.015	2.571	2.752	3.360	4.032	4.773
6	.718	.908	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.315
7	.711	.896	1.119	1.415	1.895	2.357	2.517	2.996	3.499	4.029
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833
9	.703	.883	1.100	1.383	1.833	2.282	2.398	2.821	3.250	3.696
10	.700	.879	1.093	1.372	1.812	2.258	2.359	2.784	3.199	3.581
11	.697	.876	1.088	1.363	1.796	2.234	2.328	2.719	3.166	3.492
12	.695	.873	1.083	1.356	1.782	2.219	2.303	2.681	3.135	3.428
13	.694	.870	1.079	1.350	1.771	2.200	2.282	2.650	3.102	3.372
14	.692	.868	1.076	1.345	1.761	2.185	2.264	2.624	2.977	3.326
15	.691	.866	1.074	1.341	1.753	2.171	2.249	2.602	2.947	3.286
16	.690	.865	1.071	1.337	1.746	2.159	2.235	2.583	2.921	3.252
17	.689	.863	1.069	1.333	1.740	2.149	2.224	2.567	2.898	3.222
18	.688	.862	1.067	1.330	1.734	2.141	2.214	2.552	2.879	3.197
19	.688	.861	1.066	1.328	1.729	2.135	2.205	2.538	2.861	3.174
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.491	2.797	3.091
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030
40	.681	.853	1.053	1.303	1.684	2.033	2.133	2.423	2.714	2.971
50	.679	.849	1.047	1.299	1.673	2.019	2.109	2.403	2.678	2.937
60	.679	.848	1.045	1.296	1.671	2.010	2.099	2.390	2.660	2.915
80	.677	.845	1.042	1.290	1.660	1.994	2.081	2.361	2.626	2.871
100	.675	.842	1.037	1.282	1.648	1.982	2.056	2.330	2.581	2.813
∞	.674	.841	1.036	1.282	1.645	1.969	2.054	2.326	2.567	2.791

conditions for *inference with 1-sample means*

1. Randomization

Have an SRS or representative of population

2. 10% Condition

This is not as important to check for means, as sample sizes are usually very small, but it never hurts to be safe.

3. Nearly Normal Condition

Check for *one* of the following:

1. Given that population is roughly normally distributed
2. Large enough sample size ($n > 30$) – CLT
3. Check graph (**dotplot** is easy to make by hand) and check for plausible normality

WATCH OUT FOR OUTLIERS!!!

Our first problem with real “data”

Let’s pretend that the Chip’s Ahoy company claims a mean of 24 chips per cookie... do we have statistical evidence (at $\alpha = 0.05$) to doubt them?

Suppose we take a random sample of 10 of the company’s cookies, and get the following counts for # of chocolate chips:

21	28	19	19	23	18
18	19	26	17	26	27



One-sample *t*-test

μ = true mean # of chips per cookie

Ho: $\mu = 24$

Ha: $\mu < 24$

(because we probably wouldn’t be concerned if we got MORE than 24 chips per cookie)

Conditions:

- We have a random sample of the company’s cookies
- *(Since $n < 30$, we must make a graph...)*
The graph of sample data shows no outliers, so **normality should be plausible.**



One-sample *t*-test

μ = true mean # of chips per cookie

Ho: $\mu = 24$

Ha: $\mu < 24$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{21.75 - 24}{\frac{4.025}{\sqrt{12}}}$$

(fill in numbers in the formula, then just use the “t” and “p-value” from calculator)

$t = -1.93$ $p = 0.394$ $df = 11$

Since $p < \alpha$, we reject Ho.

We HAVE evidence that the mean # of chips per cookie is less than 24.

```

L1      L2      L3      1
18
19
26
19
23
18
26
27
EDIT CALC TESTS
1:Z-Test...
L1(13) 2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:Interval...
    
```

```

T-Test
Inpt: Data Stats
mu: 24
List: L1
Freq: 1
mu: #mu < mu0 > mu0
Calculate Draw
    
```

```

T-Test
mu < 24
t = -1.936220056
p = .0394737071
x = 21.75
Sx = 4.025486984
n = 12
    
```

Estimate the mean number of chocolate chips per cookie by using a 90% confidence interval

One-sample t -interval

$$\bar{x} \pm t^*_{df} \times \frac{s}{\sqrt{n}}$$

$$= 21.75 \pm (1.796) \times \frac{4.025}{\sqrt{12}}$$

(again, fill in numbers in the formula, then just get the interval from the calculator)

(19.633, 23.837)

We are 90% confident that the true MEAN number of chips per cookie is between 19.633 and 23.837.

```
EDIT CALC TESTS
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
8:TInterval...
```

```
TInterval
Inpt:Data Stats
x̄:21.75
Sx:4.025486983...
n:12
C-Level:.9
Calculate
```

```
TInterval
(19.663, 23.837)
x̄=21.75
Sx=4.025486984
n=12
```

Homework #14

Ch 23 p.541 #2, 7, 10, 11, 13a-c, 15, 21, 31
skip 2c & 2d