Module 10b – Inference for Mean Comparisons (continued)

Review Against All Odds: Unit 27 (Inference)

Stat Procedure Diagram – Where we are

		Descriptive Statistics (Describing Pops or Samples)		**Inferential Statistics (from Samples)**
	Variable Types	Display	Describe	Estimation
Univariate	categorical (nominal or ordinal)*	Bar Graph/Pie Chart	Counts/Percentages	*When binary/dichotomous:* Confidence interval for proportions
	quantitative/continuous	Histogram/Stem & Leaf Box Plot	Mean/St Dev (normal) Median/Min, Q1, Q3, Max (skewed)	Confidence interval for means
		Display	Describe	Significance Tests/Hypothesis Tests
Bivariate	2 categorical	Tables or Bar Graphs	Two-way tables/Crosstabulation	Chi-square test (for goodness of fit)
	1 categorical, 1 quant.	Bar Graphs	Comparison of means/averages	T-test (one sample/group, two samples/groups) ANOVA (two or more samples/groups)
	2 quant.	Scatterplot	Correlation Coef. (Coef. of determ)/ Regression Line	T-test for correlation
		Display	Describe	Significance Tests/Hypothesis Tests
Multivariate	Response Variable is Quant.	-	Ordinary Least Squares Regression (OLS)	F-test for overall model T-tests for each explanatory variable
	Response Variable is categorical (dichotomous)	-	Logistic Regression	Chi-square tests of significance

NOTE: Items highlighted in yellow are covered in this course.

**When a categorical variable has two categories, it is called dichotomous.*

Still focused on…
…***quantitative/continuous variables (inference)***

- Mean Comparisons (with ***unknown*** population standard deviation)

- Types (all are categorical IV and quantitative/continuous DV)
  - **One sample/group comparison**
  - Two sample/group comparison (matched pairs/before & after)
  - Two sample/group comparison (independent samples)
  - Two or more sample/group comparison (ANOVA)

One sample t-test (practice)

Taken from Example 20.3 on page 462:

Bacteria content in Ohio State Park swimming areas

We randomly select 20 state park swimming areas and test for coliform levels.

Assume we know that the distribution of bacteria content is relatively normally distributed but we don't know that population standard deviation so we substitute the **sample standard deviation,** which is given as **1,038 bacteria per sample.**

We calculate the mean level of bacteria from the **sample of 20** to be **1,231 bacteria per sample**. The safe level is **400 bacterial per sample** so we will test to see if our sample provides evidence that the average exceeds this standard.

One sample t-test (practice)

- Bacterial Threshold for safe swimming: **400 bacteria**
- Sample standard deviation: **1,038 bacteria** (we have to use t-test)
- Avg bacteria level for a random sample from **20** swimming locations: **1,231 bacteria**

- **Null Hyp. H_0**: State Park average = 400 bacteria (no difference)
- **Alt Hyp. H_a:** State Park average > 400 bacteria (1-sided hypothesis)

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

μ=pop mean/some standard value
$\bar{x}$=sample mean
n=sample size
s = sample standard deviation

T-test (assumptions)

A t-test requires making three assumptions.

1. The data are independent observations from a ***simple random sample.***

2. The distribution of the underlying population is ***relatively normal.***

3. We do NOT know the population ***standard deviation*** (so we estimate it based on the sample standard deviation)

Inference in practice – one sample t-test

- t =[1,231-400] / [1,038 / √20]
- t = 3.58 (the sample avg for OH State Park bacteria levels is 3.58 stand. error terms above the standard of 400 bacteria)

- We reject the null hypothesis if
  - …the absolute value of ***"t" (or test statistic)*** is greater than the ***critical value.***
  - …because if "t" is greater than the critical value, that means the ***p-value*** is less than the ***alpha level.***

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

μ=pop mean/some standard value
$\bar{x}$=sample mean
n=sample size
s = sample standard deviation

Inference in practice – one sample t-test

- t = 3.58
- How do we find the critical value?
  - We will assume an **alpha level** of .05 (5%)
  - Our **degrees of freedom** (df) are n-1: 20-1=19
  - From this, we can establish are critical value for our t-statistic

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
One-sided P	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
Two-sided P	.50	.40	.30	.20	.10	.05	.04	.02	.01	.005	.002	.001

Inference in practice – one sample t-test

- t = 3.58
- How do we find the critical value?
  - The critical value for a one-sided hypothesis at alpha .05 is 1.729.
  - Because 3.58 > 1.729 we know that the p-value is less than .05, and therefore ***we reject the null.***

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
One-sided *P*	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
Two-sided *P*	.50	.40	.30	.20	.10	.05	.04	.02	.01	.005	.002	.001

Significance test results

for t-test if alpha=.05 and df = 19