Chapter 7 - Hypothesis Tests Applied to Means

7.1 Distribution of 100 random numbers:

DV

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	7	7.0	7.0	7.0
	1.00	9	9.0	9.0	16.0
	2.00	14	14.0	14.0	30.0
	3.00	9	9.0	9.0	39.0
	4.00	16	16.0	16.0	55.0
	5.00	4	4.0	4.0	59.0
	6.00	10	10.0	10.0	69.0
	7.00	14	14.0	14.0	83.0
	8.00	13	13.0	13.0	96.0
	9.00	4	4.0	4.0	100.0
	Total	100	100.0	100.0	

mean(dv) = 4.46
st. dev(dv) = 2.687
var(dv) = 7.22

7.3 Does the Central Limit Theorem work?

The mean and standard deviation of the sample are 4.46 and 2.69. The mean and standard deviation are very close to the other parameters of the population from which the sample was drawn (4.5 and 2.7, respectively.) The mean of the distribution of means is 4.45, which is close to the population mean, and the standard deviation is 1.20.

Population Parameters	Predictions from Central Limit Theorem	Empirical Sampling distribution
$\mu = 4.5$	$\bar{X} = 4.5$	$\bar{X} = 4.45$
$\sigma^2 = 7.22$	$s^2 = \frac{\sigma^2}{n} = \frac{7.22}{5} = 1.44$	$s^2 = 1.44$

The mean of the sampling distribution is approximately correct compared to that predicted by the Central Limit theorem. The variance of the sampling distribution is almost exactly what we would have predicted..

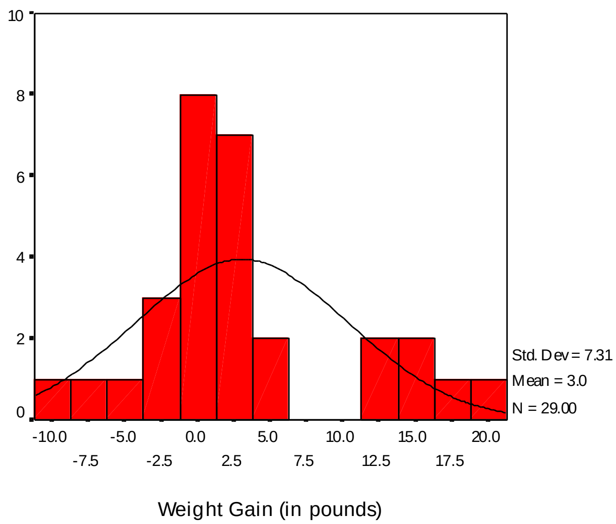
7.5 The standard error would have been smaller, because it would be estimated by $\sqrt{\frac{7.29}{15}}$ instead of $\sqrt{\frac{7.29}{5}}$.

7.7 I used a two-tailed test in the last problem, but a one-tailed test could be justified on the grounds that we had no interest is showing that these students thought that they were below average, but only in showing that they thought that they were above average.

7.9 While the group that was near the bottom certainly had less room to underestimate their performance than to overestimate it, the fact that they overestimated by so much is significant. (If they were in the bottom quartile the best that they could have scored was at the 25th percentile, yet their mean estimate was at the 68th percentile.)

7.11 Everitt's data on weight gain:

The Mean gain = 3.01, standard deviation = 7.31. t = 2.22. With 28 *df* the critical value = 2.048, so we will reject the null hypothesis and conclude that the girls gained at better than chance levels. The effect size is 3.01/7.31 = 0.41.



7.13 **a.** Performance when not reading passage

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$= \frac{46.6 - 20.0}{\frac{6.8}{\sqrt{28}}} = \frac{26.6}{1.285} = 20.70$$

b. This does not mean that the SAT is not a valid measure, but it does show that people who do well at guessing at answers also do well on the SAT. This is not very surprising.

7.15 Confidence limits on μ for Exercise 7.14:

$$CI_{.95} = \bar{X} \pm t_{.05}\frac{s}{\sqrt{n}}$$
$$= 4.39 \pm 2.03\frac{2.61}{\sqrt{36}} = 4.39 \pm 0.883$$
$$= 3.507 \le \mu \le 5.273$$

An interval formed as this one was has a probability of .95 of encompassing the mean of the population. Since this interval includes the hypothesized population mean of 3.87, it is consistent with the results in Exercise 7.14.

7.17 Confidence limits on beta-endorphin changes:

$$CI_{.95} = \bar{D} \pm t_{.05}\frac{s_D}{\sqrt{n}}$$
$$= 7.70 \pm 2.101\frac{9.945}{\sqrt{19}} = 7.70 \pm 4.794$$
$$= 2.906 \le \mu \le 12.494$$

7.19 Paired *t* test on marital satisfaction:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{D}}{s_{\bar{D}}} = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}$$
$$= \frac{2.725 - 2.791}{\frac{1.30}{\sqrt{91}}} = \frac{-.066}{.136} = -.485$$

We cannot reject the null hypothesis that males and females are equally satisfied. A paired-*t* is appropriate because it would not seem reasonable to assume that the sexual satisfaction of a husband is independent of that of his wife.

7.21 Correlation between husbands and wives:

$$r = \frac{\text{cov}_{XY}}{s_X s_Y} = \frac{0.420}{\sqrt{(1.357)(1.167)}} = \frac{0.420}{1.584} = \frac{.420}{1.259} = .334$$

The correlation between the scores of husbands and wives was .334, which is significant, and which confirms the assumption that the scores would be related.

7.23 The important question is what would the sampling distribution of the mean (or differences between means) look like, and with 91 pairs of scores that sampling distribution would be substantially continuous with a normal distribution of means.

7.25 Sullivan and Bybee study:

$$\bar{X}_{\text{int}} = 5.03 \quad s_{\text{int}} = 1.01 \quad n_{\text{int}} = 135$$
$$\bar{X}_{ctrl} = 4.61 \quad s_{\text{int}} = 1.13 \quad n_{\text{int}} = 130$$

$$t = \frac{\bar{X}_{\text{int}} - \bar{X}_{ctrl}}{\sqrt{\frac{s^2_{\text{int}}}{n_{\text{int}}} + \frac{s^2_{ctrl}}{n_{ctrl}}}} = \frac{5.03 - 4.61}{\sqrt{\frac{1.01^2}{135} + \frac{1.13^2}{130}}}$$
$$= \frac{5.03 - 4.61}{\sqrt{\frac{1.02}{135} + \frac{1.277}{130}}} = \frac{0.42}{\sqrt{0.017}} = \frac{0.42}{0.132} = 3.186$$
$$p(t > abs(3.186)) = .002$$

The quality of life was significantly better for the intervention group.

7.27 Paired *t*-test on before and after intervention quality of life

$$\bar{X}_{before} = 4.47 \quad \bar{X}_{after} = 5.03 \quad s_{diff} = 1.30 \quad n = 135$$
$$t = \frac{\bar{D} - 0}{\frac{s_{diff}}{\sqrt{n}}} = \frac{5.03 - 4.47}{\frac{1.30}{\sqrt{135}}} = \frac{0.56}{.112} = 5.00$$
$$p < .000$$

Confidence limits change in quality of life in Cognitive Behavior Therapy group:

$$CI_{.95} = \bar{D} \pm t_{.025(28)} s_{\bar{D}}$$
$$= 0.56 \pm (1.984)(0.112) = 0.56 \pm 0.222$$
$$0.34 \le \mu \le 0.78$$

The probability is .95 that this procedure has resulted in limits that bracket the mean weight gain in the population.

7.29 Katz et al (1990) study

a. Null hypothesis—there is not a significant difference in test scores between those who have read the passage and those who have not.

b. Alternative hypothesis—there is a significant difference between the two conditions.

c.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{16(10.6^2) + 27(6.8^2)}{17 + 28 - 2} = \frac{3046.24}{43} = 70.843$$

$$t = \frac{69.6 - 46.6}{\sqrt{\frac{70.843}{17} + \frac{70.843}{28}}} = \frac{23.0}{\sqrt{70.843\left(\frac{1}{17} + \frac{1}{28}\right)}} = \frac{23.0}{\sqrt{6.697}} = 8.89$$

t = 8.89 on 43 *df* if we pool the variances. This difference is significant.

d. We can conclude that students do better on this test if they read the passage on which they are going to answer questions.

7.31

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{25(63.82) + 28(53.41)}{26 + 29 - 2} = \frac{3090.98}{53} = 58.32$$

$$t = \frac{-0.45 - 3.01}{\sqrt{\frac{58.32}{26} + \frac{58.32}{29}}} = \frac{-3.46}{\sqrt{58.32\left(\frac{1}{26} + \frac{1}{29}\right)}} = \frac{-3.46}{\sqrt{4.254}} = \frac{--3.46}{2.062} = -1.68$$

A *t* on two independent groups = -1.68 on 53 *df*, which is not significant. Cognitive behavior therapy did not lead to significantly greater weight gain than the Control condition. (Variances were homogeneous.)

7.33 If those means had actually come from independent samples, we could not remove differences due to couples, and the resulting *t* would have been somewhat smaller.

7.35 The difference between the two answers in not greater than it is because the correlation between husbands and wives was actually quite low.

7.37 **a.** I would assume that the experimental hypothesis is the hypothesis that mothers of schizophrenic children provide TAT descriptions that show less positive parent-child relationships.

b.

Normal	Mean = 3.55	$s = 1.887$	$n = 20$
Schizophrenic	Mean = 2.10	$s = 1.553$	$n = 20$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{3.55 - 2.10}{\sqrt{\frac{1.887^2}{20} + \frac{1.553^2}{20}}}$$

$$= \frac{1.45}{\sqrt{0.299}} = \frac{1.45}{0.546} = 2.66$$

[$t_{.05}(38) = \pm 2.02$] Reject the null hypothesis

This t is significant on 38 *df*, and I would conclude that the mean number of pictures portraying positive parent-child relationships is lower in the schizophrenic group than in the normal group.

7.39 There is no way to tell cause and effect relationships in Exercise 7.37. It could be that people who experience poor parent-child interaction are at risk for schizophrenia. But it could also be that schizophrenic children disrupt the family and poor relationships come as a result.

7.41 95% confidence limits:

$$CI_{.05} = \left(\bar{X}_1 - \bar{X}_2\right) \pm t_{.025}\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$$

$$= \left(18.778 - 17.625\right) \pm (2.131)\sqrt{\frac{16.362}{9} + \frac{16.362}{8}} = 1.153 \pm 4.189$$

$$-3.036 \le \left(\mu_1 - \mu_2\right) \le 5.342$$

7.43 Repeating Exercise 7.42 with time as the dependent variable:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{2.102 - 1.246}{\sqrt{\frac{0.714}{5} + \frac{0.091}{5}}} = \frac{0.856}{\sqrt{0.161}} = \frac{0.856}{0.401} = 2.134$$

The variances are very different, but even if we did not adjust the degrees of freedom, we would still fail to reject the null hypothesis.

7.45 If you take the absolute differences between the observations and their group means and run a *t* test comparing the two groups on the absolute differences, you obtain $t = 0.625$. Squaring this you have $F = 0.391$, which makes it clear that Levene's test in SPSS is operating on the absolute differences. (The *t* for squared differences would equal 0.213, which would give an *F* of 0.045.)

7.47 Differences between males and females on anxiety and depression:

(We cannot assume homogeneity of regression here.)

Independent Samples Test

Equal variances not assumed

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DEPRESST	3.256	248.346	.001	3.426	1.052	1.353	5.499
ANXT	1.670	246.260	.096	1.805	1.081	-.324	3.933

7.49 Effect size for data in Exercise 7.27:

$$d = \frac{\bar{X}_{After} - \bar{X}_{Before}}{s_{Before}} = \frac{0.56}{1.18} = 0.47$$

I chose to use the standard deviation of the before therapy scores because it provides a reasonable base against which to standardize the mean difference. The confidence intervals on the difference, which is another way to examine the size of an effect, were given in the answer to

Exercise 7.27. (Katherine Long at Fordham pointed me to this error and to the errors in 7.25 and 7.27. This question really refers back to 7.27.)

7.50 Effect size for data in Exercise 7.31

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p} = \frac{-0.45 - 7.26}{\sqrt{58.91}} = \frac{-7.71}{7.681} = 1.00$$

The two means are approximately 1 standard deviation apart.

7.51 **a.** The scale of measurement is important because if we rescaled the categories as 1, 2, 4, and 6, for example, we would have quite different answers.

b. The first exercise asks if there is a relationship between the satisfaction of husbands and wives. The second simply asks if males (husbands) are more satisfied, on average, than females (wives).

c. You could adapt the suggestion made in the text about combining the *t* on independent groups and the *t* on matched groups.

d. I'm really not very comfortable with the *t* test because I am not pleased with the scale of measurement. An alternative would be a ranked test, but the number of ties is huge, and that probably worries me even more.