Final Exam

Statistics 1XXX

1. *Confidence interval for a population mean μ. This is a z-CI problem. For t-CI problems please see your notes and quiz.* A sample of 49 measurements of tensile strength (roof hanger) are calculated to have a mean of 2.45 and a standard deviation of 0.25. Determine the 95% confidence interval for the measurements of all hangers.

   Solution. The 95% confidence interval for a population mean (the sample size is larger than 30) is given by
$$\bar{X} \pm z_{\alpha/2}\frac{s}{\sqrt{n}},$$
   that is,
$$2.45 \pm 1.96\frac{.25}{\sqrt{49}} = 2.45 \pm .07$$
   or, in form of interval,
$$[2.38; 2.52].$$

2. *Confidence interval for a population proportion p.* In a survey two days before election day, the Democratic candidate for the Senate was supported by 918 voters. His opponent was supported by 582 voters. Construct 99% confidence interval for the proportion of all voters who support the Democratic candidate.

   Solution. The 99% confidence interval for a population proportion (both $n\hat{p}$ and $n\hat{p}(1-\hat{p})$ are greater than 5) is given by
$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$
   that is,
$$\frac{918}{1500} \pm 2.576\sqrt{\frac{\frac{918}{1500}\frac{582}{1500}}{1500}} = .612 \pm .032$$
   or, in form of interval,
$$[.58; .644].$$

3. *Sample size determination. This is a mean problem. For proportion problems please see your notes and quiz.* A statistician wants to estimate the mean weekly family expenditure on clothes. He believes that the standard deviation is about \$125. Determine with 99% confidence the number of families that must be sampled to estimate the mean weekly family expenditure on clothes to within ±\$15.
   Solution.
   The required sample size:
$$n \simeq \left(\frac{z_{\alpha/2}s}{W}\right)^2 = \frac{2.576^2 \times 125^2}{15^2} \simeq 461.$$

4. *z-test about a population mean μ.* The air pumps at service stations come equipped with a gauge to regulate the air pressure of tires. A mechanic believes that the gauges are in error by at least 3 pounds per square inch. To test his belief he takes a random sample of 50 air pump gauges and determines the absolute difference between the true pressure (as measured by an accurate measuring device) and the pressure shown on the air pump gauge. The mean and the standard deviation of the sample are $\bar{X} = 3.4$ and $s = 1.2$. Can the mechanic infer that he is correct at the 5% significance level?

Solution.
1. **Setup**
$H_0 : \mu = 3$
$H_a : \mu > 3$
2. **Test statistic** (the sample size is larger than 30)

$$z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{3.4 - 3}{1.2/\sqrt{50}} = 2.36$$

3. **Rejection region**

$$RR = [z_\alpha; +\infty) = [z_{.05}; +\infty) = [1.645; +\infty)$$

4. **Conclusion**
Since the test statistic is in the rejection region we accept H_a at 5% significance level. That is, the data support the mechanic's believe that the gauge are in error by at least 3 pounds per square inch.

5. *t-test about a population mean μ.* During the past month Debra stopped at the Soaring Eagle Casino in Mount Pleasant, Michigan, 6 times with no intention to play (since she has no money). Having nothing better to do, she counted how many jackpots the players received. The results follow:

   14 11 20 15 12 24

   Assume that the number of jackpots is normally distributed. Can she infer at the 5% significance level that the population average number of jackpots is over 15?

Solution.
1. **Setup**
$H_0 : \mu = 15$
$H_a : \mu > 15$
2. **Test statistic** (small sample size, normally distributed observations)

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}.$$

Since

$$\bar{X} = (14 + 11 + 20 + 15 + 12 + 24)/6 = 16$$

and

$$s = \sqrt{\left((14-16)^2 + (11-16)^2 + (20-16)^2 + (15-16)^2 + (12-16)^2 + (24-16)^2\right)/5}$$

$$= 5.02,$$

we get

$$t = \frac{16 - 15}{5.02/\sqrt{6}} = 0.49$$

3. **Rejection region**

$$RR = [t_{\alpha,n-1}; +\infty) = [t_{.05,5}; +\infty) = [2.02; +\infty)$$

4. **Conclusion**
Since the test statistic does not fall into the rejection region we reject H_a at 5% significance level. That is, the data do not support the Debra's believe that the average number of jackpots is over 15.

6. *z-test about a population proportion p.* In clinical studies of an allergy drug, 81 of the 900 subjects experienced drowsiness. A competitor claims that at least 10% of the users of this drug experience drowsiness.
   a. Is there enough evidence at the 5% significance level to infer that the competitor is correct?
   b. Compute the p-value of the test in the part (a).

   Solution
   a.
   1. **Setup**
   $H_0 : p = .10$
   $H_a : p > .10$
   2. **Test statistic** $(np_0(1 - p_0) > 5)$

   $$z = \frac{\hat{p} - p_0}{\sqrt{(p_0(1 - p_0))/n}} = \frac{81/900 - .1}{\sqrt{.1 \times .9/900}} = -1$$

   3. **Rejection region**

   $$RR = [z_\alpha; +\infty) = [z_{.05}; +\infty) = [1.645; +\infty)$$

   4. **Conclusion**
   Since the test statistic does not fall into the rejection region we reject H_a at 5% significance level. That is, the data do not support the competitor's claim that 10% of users of this drug experience drowsiness.

   b.
   p-value= $P(z > -1) = .84 > .05$. Since the p-value is larger than the significance level we reject H_a.

7. *Sampling distribution for $\bar{X}$.* Suppose that a random sample of size 64 is to be selected from a population with mean 40 and standard deviation 5.
   a. What is the mean and standard deviation of sample mean $\bar{X}$?
   b. What is the probability that $\bar{X}$ will be greater than 40.7?

   Solution
   a. The mean of $\bar{X}$ is 40, and the standard deviation is $5/\sqrt{64} = .625$.
   b. Since the sample size is above 30, approximately $\bar{X} \sim N(40, .625^2)$. Therefore,

   $$\begin{aligned} P(\bar{X} > 40.7) &= P\left(Z > \frac{40.7 - 40}{.625}\right) \\ &= P(Z > 1.12) \\ &= 1 - P(Z < 1.12) \\ &= 1 - .8686 = .1314 \end{aligned}$$

8. *Sampling distribution for $\hat{p}$.* A random sample is to be selected from a population that has a proportion of successes $p = 0.65$.
   a. Determine the mean and standard deviation of the sample proportion $\hat{p}$ for the sample size $n = 100$.
   b. When $n = 100$, what is $P(0.6 < \hat{p} < 0.7)$?

   Solution
   a. The mean of $\hat{p}$ is .65, and the standard deviation is $\sqrt{.65 \times .35/100} \approx .0477$.
   b. Since the sample size is large $(np(1 - p) > 5)$, approximately $\hat{p} \sim N(.65, .0477^2)$. Therefore,

   $$\begin{aligned} P(.6 < \hat{p} < .7) &= P\left(\frac{.6 - .65}{.0477} < Z < \frac{.7 - .65}{.0477}\right) \\ &= P(-1.05 < Z < 1.05) \\ &= P(Z < 1.05) - P(Z < -1.05) \\ &= .8531 - .1469 = .7062 \end{aligned}$$

9. *Comparing two population means: independent samples.* Do government employees take longer coffee breaks than private sector workers? That is a question that interested a management consultant. To examine the issue, he took a random sample of ten government employees and another random sample of ten private sector workers and measured the amount of time (in minutes) they spent in coffee breaks during the day. The results are listed below. Do these data provide sufficient evidence at the 5% significance level to support the consultants claim?

Private Sectors Workers	Government Employees
25	23
19	18
18	34
22	31
28	28
25	33
21	25
21	27
20	32
16	21

Note:
$\bar{X}_1 = 21.50,\ s_1 = 3.63;\ \bar{X}_2 = 27.20,\ s_2 = 5.41;$
$\bar{X}_D = 5.70,\ s_D = 5.79.$
Solution.
1. **Setup**
$H_0 : \mu_1 = \mu_2$
$H_a : \mu_1 < \mu_2$
Here μ_2 is the average amount of time spent in coffee breaks during the day by government employees, and μ_1 is the average amount of time the private sector workers spend in coffee breaks.
2. **Test statistic** (2-sample t-test)

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

$$= \frac{21.50 - 27.20}{\sqrt{3.63^2/10 + 5.41^2/10}} = -2.766$$

3. **Rejection Region**
In this case we have independent samples, small sample sizes, one-tailed alternative, and $\nu = 16$, so the rejection region is

$$RR = (-\infty; -t_{\alpha,\nu}] = (-\infty; -t_{.05,16}] = (-\infty; -1.746].$$

4. **Conclusion**
Since the test statistic falls into the rejection region we reject H_0 at 5% significance level in favor of the alternative. That is, the data support the consultant's claim that government employees take longer breaks than private sector workers.

10. *Comparing two population means: paired test.* A marketing consultant was in the process of studying the perception of married couples concerning their weekly food expenditures. He believed that the husbands perception would be higher than the wifes. To judge his belief, he takes a random sample of ten married couples and asks each spouse to estimate the family food expenditure (in dollars) during the previous week. The data are shown below. Can the consultant conclude at the 5% significance level that the husbands estimate is higher than the wifes estimate?

Husband	Wife
380	270
280	300
215	185
350	320
210	180
410	390
250	250
360	320
180	170
400	330

Note:
$\bar{X}_1 = 303.5$, $s_1 = 86.3$; $\bar{X}_2 = 271.5$, $s_2 = 74.2$;
$\bar{X}_D = 32.0$, $s_D = 36.5$.
Solution.
1. **Setup**
$H_0 : \mu_D = 0$
$H_a : \mu_D > 0$
where μ_D is the average difference between the husband's estimate and his wife's estimate.
2. **Test statistic** (paired t-test)

$$t = \frac{\bar{X}_D - \mu_0}{s_D/\sqrt{n}} = \frac{32}{36.5/\sqrt{10}} = 2.776$$

3. **Rejection Region**
In this case we have paired observations, small sample size, and one-tailed alternative, so the rejection region is

$$RR = [t_{\alpha,n-1}; +\infty) = [t_{.05,9}; +\infty) = [1.833; +\infty).$$

4. **Conclusion**
Since the test statistic falls into the rejection region we reject H_0 at 5% significance level in favor of the alternative. That is, the data support the marketing consultant's claim that the husbands' estimate is higher than the wife estimate.