Economics 154-257

V. Zinde-Walsh

Topic 1. **Hypotheses testing.**

Lecture 1.

Introduction.

So far we have **estimated parameters** of some distributions: μ and σ^2 for the normal, $N(\mu, \sigma^2)$, by $\bar{X}$ and s^2; p (proportion) for the Bernoulli by $\hat{p}$; estimating these parameters provided an estimated distribution, when we assumed that the shape of the distribution (e.g. normal) was known.

In some cases we may not know the shape of the distribution but are anyway only interested in some **moments**, say, the mean, and estimated it by $\bar{X}$.

We have learned about **properties of estimators** that are useful in judging how well we are estimating: consistency, unbiasedness, efficiency. E.g. we know that estimating the mean by sample average can give a consistent and unbiased estimator that is also efficient (BLUE).

Since the estimators are random we worked on establishing their distributions, called **sampling distributions** and found that for some we can find exact distributions, such as χ^2 for scaled s^2 for random sample from normal distribution. Distribution for some estimators can be **approximated** by known distributions, e.g. normal; under suitable conditions $\bar{X}$ or $\hat{p}$ are asymptotically normally distributed.

Recognizing that a point estimator such as $\bar{X}$ cannot give the parameter value precisely but only with some margin of error, we constructed **confidence intervals** to reflect that. Confidence intervals tell us that we have 1-α confidence (say, 95%) of the population parameter being within the bounds of the interval.

We may have **specific questions** of interest regarding the estimated parameter.

Example.

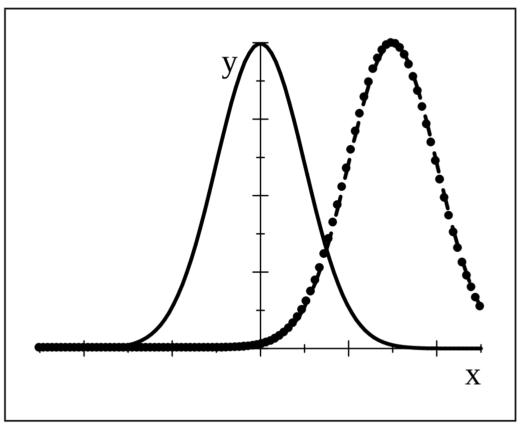
A supplier usually sends chips with power usage rate of 90 milliamps, but sometimes does not have enough inventory and substitutes chips with 120 milliamps power usage rate; if you use these inferior chips you need to implement expensive procedures to reduce overall power usage. You receive a supply and wish to determine whether the chips have power usage rate of 90 or 120. Your testing procedure destroys the tested chips so you wish to limit that expense as well (so, e.g. test a small sample of he chips in the shipment).

Questions: Did the chips shipment come from the usual batch with power rate of 90? How likely is it that the chips came from the batch with the rate of 90?

Formal question in the form of a **hypothesis.**

We assume the actual power usage rate of the individual chips in the shipment is random, and has a distribution with mean either 90 or 120.

In the example two possibilities, or two "states of the world". One is $\mu = 90$, the other $\mu = 120$.



Call one the **null hypothesis**: $H_0 : \mu = 90$; the other the **alternative hypothesis** $H_1 : \mu = 120$.

Decision consists in picking one of the two; a reasonable **decision rule** is needed.

Types of decision **error.**

		decision mean=90	decision mean=120
state of the world	mean=90	correct	Type I error
	mean=120	Type II error	correct

So **type I error** is when we reject the true null-hypothesis: decide that mean is 120 when it is 90 as in H_o.

Type II error is when we incorrectly accept the null: decide that 90 (H_0) is right when it is not (120).

To make a decision we rely on **sample information**. For example, we compute the sample mean, $\bar{X}$, and base our decision on its value.

How is this done? by deciding on a critical or **rejection region** bounded by, say, b and the **rule**:

If $\bar{X} > b$ reject H_o in favor of H_1.

Each decision rule is associated with possibility of errors.

Under some conditions we can find **probabilities** for type I and II errors.

Computation of errors associated with various decision rules for the example.

Assume that the power rates are normally distributed $N(\mu, \sigma^2)$ with unknown mean μ (either 90 or 120) and $\sigma^2 = 3600$, so that st.deviation is 60. Take a sample of size $n = 25$; then $\bar{X}$ has distribution $N(\mu, 144)$ (where $144 = \frac{3600}{25}$).

Candidate decision rule 1:

Take the mid-point between 90 and 120, 105 and decide on H_0 if $\bar{X} < 105$, H_1 otherwise.

Then $\alpha = \text{Pr}(Type \text{ I error})$

$=\text{Pr}(\bar{X} > 105 | N(90, 144)$

$= \text{Pr}(Z > \frac{105-90}{12}) = \text{Pr}(Z > 1.25) = .106;$

$\beta = \text{Pr}(Type \text{ II error})$

$=\text{Pr}(\bar{X} < 105 | N(120, 144) = \text{Pr}(Z < \frac{105-120}{12}) = \text{Pr}(Z < -1.25) = .106.$

What happens if decision rule changes?

Decision rule 2:

Decide on H_0 if $\bar{X} < 110$, H_1 otherwise.

Then $\alpha = \text{Pr}(Type \text{ I error})$

$=\text{Pr}(\bar{X} > 110 | N(90, 144) = \text{Pr}(Z > \frac{110-90}{12})$

$= \Pr(Z > 1.6667) = .0475;$

$\beta = \Pr(Type\ \text{II error})$

$= \Pr(\bar{X} < 110 | N(120, 144) = \Pr(Z < \frac{110-120}{12}) = \Pr(Z < -0.83333) = .2,$

so probability of Type I error, α, went down, but for Type II error, β, went up.

Probability $1 - \beta$ is called the **power of the test** (the probability of a correct rejection of the null).

		decision	
		accept H_0	reject H_o
state of the world	H_o	correct, $\Pr = 1 - \alpha$	Pr(T I error)$=\alpha$
	H_1	Pr(T II error)$=\beta$	correct, $\Pr = 1 - \beta$

There is a **trade-off** between α and β.

Some other aspects of the problem have an effect.

Both α and β decline with the following changes.

If the null and alternative are farther apart: e.g. $H_0 : \mu = 90$, and $H_1 : \mu = 150$.

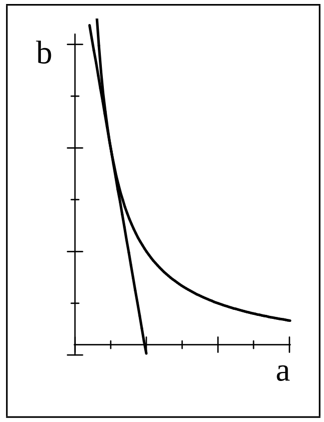
If the variance is smaller, e.g. $\sigma^2 = 3000$, rather than 3600.

If the sample size is larger, e.g. $n = 50$ instead of 25.

Decision rule.

How to choose a reasonable decision rule?

Type I and type II errors are usually costly, each with its own cost. Suppose that we can find the trade-off between the two (e.g. in the form of relative price) of α and β. Then if we could plot β against α we could find the point of tangency that would minimize the cost of the error.



Usually this is not done. The problem is that as we shall later see in a test of composite hypotheses (with numerous alternatives) the value of β is not unique and depends on which specific alternative is true, so that evaluating the cost of type II error is not straightforward. This is why the focus is on controlling α.

Decision rule is often based on choosing some small **significance level** α (=Pr($type\ I$ error), such as .05, or .01. Historically this had to do with statistical evaluation of agricultural experiments for effectiveness of innovations: new seeds, fertilizes, etc. The idea was not to introduce a new product unless the evidence for its effectiveness was very clear, so that type I error would be small. Similarly for medical innovations. Or for new economic theories.

Since there is no symmetry between the null and alternative it is important to give some thought to setting up the hypothesis testing problem.

The hypotheses that we considered in the example are called **simple:** a point null against a point alternative; in this case both distributions (for H_0 and H_1) are completely specified.

Composite hypotheses.

Now we specify a whole range of values for the alternative.

Example $H_o : \mu = 90$, but now the alternative is $H_1 : \mu > 90$.

Generally,

$H_o : \mu = a$; $H_1 : \mu > a$

or

$H_o : \mu = a$; $H_1 : \mu < a$

is called a **one-sided test**.

Sometimes the null H_0 in a one-sided test is represented by $\leq$ or $\geq$, e.g. $H_0 : \mu \leq a$ versus $H_1 : \mu > a$. This makes sense because if the null $\mu = a$ is rejected in favor of $\mu > a$, then any $\mu < a$ is automatically rejected as well.

We shall also consider **two-sided tests**:

$H_o : \mu = a$; $H_1 : \mu \neq a$.

Tests of a population mean in a normal population (population variance known).

Let X be $N\left(\mu, \sigma^2\right)$ where σ^2 is known. The hypothesis is about the population mean, μ, $H_0 : \mu = a$.

Any evidence about the value of the parameter μ comes from sample data, and we usually use the sample mean $\bar{X}$ (random) as an estimator of μ. Using the sampling distribution of the sample mean we can standardize and obtain the corresponding standard normal $N(0,1)$ value

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}.$$

Notice, that given the null hypothesis $H_0 : \mu = a$, we can actually compute the value of the statistic Z for the null distribution.

For any $\alpha-$ level of the test, the critical rejection region can be found by finding the corresponding critical value in the normal table, e.g. $Z_{.025} = 1.96$.

Then comparing the computed value of the statistic with the critical we can see whether it falls into the rejection region.

Example.

For the variable X distributed $N(\mu, \sigma^2)$, where $\sigma^2 =$ 3600 test the null hypothesis

$H_o : \mu = 90$;

against the alternative

$H_1 : \mu > 90$

based on a sample of size $n = 25$ that provided a sample mean of $\bar{X} = 105$.

The distribution of $\bar{X}$ under H_0 is $N(\mu, \frac{\sigma^2}{n})$, here $N(90, 144)$ where the sampling variance of $\bar{X}$ is $144 = \frac{3600}{25}$.

Consider a test at **5% significance level**, or $\alpha = .05$. Then the **decision rule** is reject if $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} > Z_\alpha$ where $\Pr(Z > Z_\alpha) = \alpha$. So we need to find **the critical value** Z_α for $\alpha = .05$; it is $Z_{.05} = 1.65$.

Now compute the **value of the test statistic**, $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$. Here it is $\frac{105-90}{12} = 1.25$.

Decision: since it is not larger than the critical value we **do not reject** H_0.

One-sided tests.

To summarize.

To test $H_o : \mu = a$; $H_1 : \mu > a$

at significance level α

(a) compute the test statistic: $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$;

(b) find the critical value: Z_α;

(c) compare the value of Z and Z_a; if $Z > Z_a$, reject H_0, if not, do not reject.

Example (continued).

Test with the same information the hypotheses $H_o : \mu = 120$; $H_1 : \mu < 120$.

To test $H_o : \mu = a$; $H_1 : \mu < a$

at significance level $\alpha = .1$

(a) compute the test statistic: $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$; here $\frac{105 - 120}{12} = -1.25$

(b) find the critical value: -Z_α; here -$Z_\alpha = -1.29$

(c) compare the value of Z and -Z_a; if $Z <$-Z_a, reject H_0, if not, do not reject. Here -1.25>-1.29, so do not reject (although it is close).

Would the null have been rejected at 5%? at 25%?

Check.

Would with the same sample information a one-sided test of $H_0 : \mu = 90$ versus $H_1 : \mu < 90$ have made sense? Explain.

Two-sided tests.

To test $H_o : \mu = a$; $H_1 : \mu \neq a$

at significance level α

(a) compute the test statistic: $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$.

(b) find the critical value: $Z_{\alpha/2}$.

(c) compare the computed value of $|Z|$ with $Z_{\alpha/2}$; if $|Z| > Z_{\alpha/2}$ reject H_0.

Example.

The average grade in a class of 100 was 67; assuming that the grades are normally distributed with population variance of 225 is it credible to say that the population mean is 70? Perform a test at 5% significance level.

$H_0 : \mu = 70; H_1 : \mu \neq 70.$

Test statistic $\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{67-70}{\sqrt{\frac{225}{100}}} = -2.0$

Critical value is $Z_{.025} = 1.96.$

The null is rejected.

What would happen in a one-sided test at $\alpha = .05$?

Explain.

Relation to confidence intervals.

We show that a $1-\alpha$ confidence interval for the population parameter θ is such that any null specifying θ within that interval cannot be rejected at α level of significance.

Example:

(a) $H_0 : \mu = 0$ in a normal distribution with known variance $\sigma^2 = 4$, then the sampling distribution of the sample mean, $\bar{X}$, under H_0 is $N(0, \frac{4}{n})$. If $\alpha = .05$ and in a sample of 100 we observe $\bar{X} = .1$.

The value of the test statistic is

$$Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{.1}{\sqrt{\frac{4}{100}}} = 0.5.$$

The critical value is 1.96; we cannot reject H_0.

(b) The 95% CI for μ is $(\bar{X}-Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \bar{X}+Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}})$.

Here

$$
\begin{aligned}
& (.1 - 1.96\sqrt{\frac{4}{100}}, .1 + 1.96\sqrt{\frac{4}{100}}) \\
= & \ (\ -0.292 \, , \ \ 0.492 \ \ ).
\end{aligned}
$$

We see that $\mu = 0$ is inside.

Theorem. (i) For any value μ_0 that is inside the 1-α (two-sided) CI the null hypothesis $H_0 : \mu = \mu_0$ cannot be rejected against a two-sided alternative $H_1 : \mu \neq \mu_0$ at α level of significance. (ii) Any value μ_0 that is not rejected at level α is inside the $(1-\alpha)$ confidence interval.

Proof.

(i) If μ_0 is inside the CI, then μ_0 is such that

$$
\bar{X} - Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} < \mu_0 < \bar{X} + Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}.
$$

Then $\left|\frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right| < Z_{\alpha/2}$, so H_0 cannot be rejected for that μ_0.

(ii) If $H_0 : \mu = \mu_0$ is not rejected vs $H_1 : \mu \neq \mu_0$ then we had $\left|\frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right| < Z_{\alpha/2}$, thus

$$\bar{X} - Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} < \mu_0 < \bar{X} + Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}$$

and μ_0 is inside the $(1-\alpha)CI.$■

The 1-α CI coincides with the set of values for which the null hypothesis cannot be rejected at significance level α.

The same is true for one-sided CI.

This implies that if you know how to contsruct CI's you can test hypotheses, if you know how to test a hypothesis you can construct a confidence interval.

Tests of a mean of a normal distribution (population variance not known).

Similarly, to construction of CI for this case the population variance is estimated by sample variance, s^2; standardizing we get the $t-$**ratio**$=\frac{\bar{X}-\mu}{s/\sqrt{n}}$ (instead of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z$, st. normal); the distribution of $t-ratio$ is **Student's** t **with** $n-1$ **d.f.**. The testing strategy is the same as with known variance, just using estimated variance and the critical values from the $t-$distribution.

So, for example for a one sided test.

To test $H_o : \mu = a$; $H_1 : \mu < a$

at significance level α

(a) compute the test statistic, $t-$ratio: $t = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}}$;

(b) find the critical value: -$t_{n-1,\alpha}$;

(c) compare the value of statistic t and -$t_{n-1,a}$; if $t <$-$t_{n-1,a}$, reject H_0, if not, do not reject.

Example.

A supplier usually sends chips with power usage rate of 90 milliamps, but sometimes does not have enough inventory and substitutes chips with a higher power usage rate. Assume that the power usage in the shipment is normally distributed. In a random sample of size $n = 25$ the sample mean was $\bar{X} = 105$, the sample variance $s^2 = 3000$. Test whether the expected power usage rate in the shipment is above 90.

What is given?

$X \sim N(\mu, \sigma^2)$, so sampling distribution of $\bar{X}$ in sample of size n is $N(\mu, \frac{\sigma^2}{n})$;

σ^2 is estimated by s^2.

Set up the hypotheses H_0, H_1.

H_0 : $\mu = 90$ (or $\mu \leq 90$); H_1 :$\mu > 90$.

For decision rule select α, level of the test.

Say, $\alpha =$ .05.

Test statistic: $t = \frac{105-90}{\sqrt{\frac{3000}{25}}} = 1.369$.

Critical value $t_{24,.05} = 1.711$.

Compare: $t - ratio < t_{24.05}$ so **cannot reject** H_0.

If tested at $\alpha =$.1 level, the critical value is $t_{24,.1} =$ 1.318, so at this level we reject.

Probability value; prob-value, p-value of a test statistic.

Since α is chosen fairly arbitrarily and we see that one gets a "weak" rejection or acceptance sometimes, meaning that the statistic is very close to a critical value, while sometimes the result is "strong", e.g. we reject at 10%

but could reject even at 1% one may ask, how can this strength of rejection or acceptance be captured.

A way to do this is to compute **p-value** of the sample statistic. Call the statistic ($t - ratio$ or some other), $\hat{S}$ (e.g., t); it has a known distribution (e.g., Student's $t_{d.f.}$), so that $\Pr(S > \hat{S})$ can be found.

Definition. For a one-sided alternative (in H_1 "$>$") $p-$value$=$ $\Pr(S > \hat{S})$ where probability is found according to the (known) distribution of the test statistic under H_0. For a two-sided alternative ($\neq$ in H_1) $p - value = 2\Pr(S > \hat{S})$.

$P-value$ is the probability of getting values of the statistic **more extreme (in the direction of the alternative)** than the value actually observed.

In the **example** above, assume for simplicity that σ^2 is known and equals 3000. Then we would have the test statistic distributed as $N(\mu, \frac{\sigma^2}{n})$; for our example under

H_0 the distribution is $N(90, \frac{3000}{25})$. The value of the statistic $\frac{105-90}{\sqrt{\frac{3000}{25}}}$ was 1.369. What is the p-value?

By definition $p - value = \Pr(Z > 1.369) = 1 - .9147 = 0.085\,3$.

What does the p-value tell us? At what levels of significance can we reject the null?

We can call this $p-$value α', in the example if α' were the level of the test, then we would have that $1.369 = Z_{\alpha'}$, the critical value for $\alpha' = 0.085\,3$. So we would be just indifferent whether to reject or accept. But if we had α such that critical $Z_a < 1.39$, we would reject.

When would Z_α be less than $Z_{\alpha'}$? When $\alpha > \alpha'$.

So in the example it would require the level to be higher than 0.085 3 to be able to reject. So we reject for all $a > 0.085\,3$, and do not reject for all $\alpha < 0.085\,3$.

Specifically, in a test at 5% ($\alpha = .05$) we would not reject, but for the test at 10% ($\alpha = .1$) level we would reject the null.

Tests of a population proportion.

$H_o : p = p_0$ against (usually) a one-sided, or a two-sided alternative.

Exact distribution. If sample size is small the sampling distribution of the sample proportion under the null hypothesis can be constructed exactly, since the sample proportion, $\hat{p} = \frac{k}{n}$, is distributed as $\frac{1}{n}$ times a binomial random variable k (with known p_0 under the null-hypothesis).

So you could do this test exactly in a small sample: use the binomial probabilities to find critical values for significance level α. Even for larger samples tables of critical values are constructed, or calculations can be performed by special programs to get the exact distribution. This is particularly important when the normal distribution does not provide a good approximation: even when sample size is large if the p is extreme: small or close to 1, np, or $n(1-p)$ may be very small ("a rule-of-thumb": if $np(1-p) < 5$, the normal is a poor approximation to binomial. in that case use the exact distribution).

Example.

Claim: the probability, p, of a long slowdown in the metro in rush-hour is at most 0.2. Taking the metro a student was late for three of her 6 morning exams in one year (since she avoids taking morning classes this was her only available info - so small sample). She observes $\hat{p} = \frac{3}{6} = 0.5$. She does not believe the claim and wants to test at 10% significance level.

Formulate: $H_0 : p \leq .2; H_1 : p > .2$.

The probability mass function for $\hat{p}$ under the null is ($\hat{p}$ takes values: $\frac{0}{6}, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, \frac{6}{6}$) :

$\mathsf{Pr}(\frac{k}{6}) = \mathsf{Pr}(k$ out of 6 in binomial$)$

$= C_k^6 p^k (1-p)^{6-k} = \frac{6!}{k!(6-k)!} p^k (1-p)^{6-k}$ for $p = .2$.

proportion of slowdowns in 6 tries	prob.
0	0.26
$\frac{1}{6}$	0.39
$\frac{2}{6}$	0.25
$\frac{3}{6}$	0.08
$\frac{4}{6}$	0.02
$\frac{5}{6}$	.002
1	≈ 0

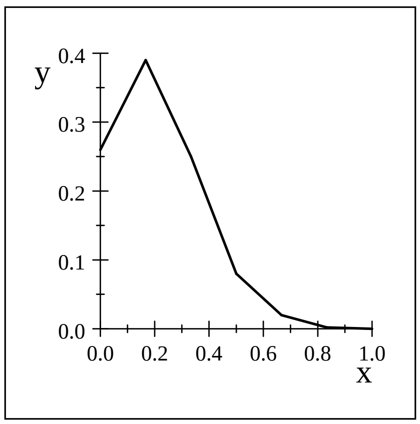
The issue of p-value for a discrete distribution.

For a continuous distribution the p-value (e.g. in a one-sided case) is $\mathsf{Pr}(\tau > \tau^*) = \mathsf{Pr}(\tau \geq \tau^*)$, where τ^* is the observed value of the statistic τ (could be standard normal, or $t_{d.f.}$).

The meaning of a $p-$value is to say if the null is true, how plausible is the value observed?

Typically the p-value for the discrete distribution is computed as Pr($\tau \geq \tau^*$).

What is the p-value for $\frac{3}{6}$? Pr $\left(p \geq \frac{3}{6}\right) = .102$, so based on this the H_0 would not be rejected at $\alpha = .1$.



Large sample size.

In a large sample the normal approximation can be used.

Example.

"The grandmother hypothesis: rocking a baby stops crying".

How to set up the test? What are the null and alternative?

If rocking has no effect on crying then we may assume that a baby stops crying randomly regardless of being rocked and will stop within some interval **with probability .5** whether it is rocked or not. The grandmother thinks that a rocked baby will stop crying with a **higher probability**.

This set-up implies a one-sided alternative:

$H_0 : p = .5$; $H_1 : p > .5$.

To test we need to collect some sample info, compute the value of a test statistic and consider the sampling distribution of the test statistic. Here, this would be the sample proportion, $\hat{p}$.

Suppose that a sample of n =126 crying babies were rocked for a while and $k = 70$ of those stopped crying within a specified period.

Then $\hat{p} = \frac{79}{126} = 0.627$.

The sampling distribution under the null is approximately normal; its mean is $p = .5$, $var(\hat{p}) = \frac{p(1-p)}{n} = \frac{.5 \cdot .5}{126} = .00198$, so $N(.5, .002)$. The value of the test statistic is (standardized)

$$Z = \frac{.627 - .5}{\sqrt{.00198}} = 2.85$$

Set the level of the test at $\alpha = .01$. The critical value $Z_{.01} = 2.33$.

We reject the null hypothesis in favor of the "grandmother hypothesis".

Suppose that you were just interested in whether there is some effect of rocking on a crying baby, regardless of whether it was to stop the baby from crying, or to increase the probability of crying. Then you would specify a two-sided alternative: $H_0 : p = .5$ (no effect) against $H_1 : p \neq .5$ (some effect, either positive, or negative).

If you wished to test at the same significance level, all that would change is the critical value: for $\alpha = .01$, the critical value now is $Z_{\alpha/2} = Z_{.005} = 2.57$. (note: this is bigger than Z_α, so at the same level it is harder to reject a two-sided hypothesis).

Probability of Type II error.

Recall that probability of Type I error$=\alpha =$ the level of the test.

Probability of Type II error$=\beta =$ Pr(H_0 not rejected$|H_1$ true).

β depends on the specific alternative and changes as a function of the true alternative.

Exercise: compute β for test of a mean.

$H_0 : \mu = 0$ in a normal distribution with known variance $\sigma^2 = 4$.

$H_1 : \mu > 0$.

The sampling distribution of the sample mean, $\bar{X}$, under H_0 is $N(0, \frac{4}{n})$.

Set $\alpha = .05$.

In a sample of 100 we observe $\bar{X} = .1$ then the standardized value of the statistic is $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{.1}{\sqrt{\frac{4}{100}}} = 0.5$.

Compare to the critical value, $Z_{.05} = 1.645$.

H_0 cannot be rejected on this evidence: $0.5 < 1.645$.

In fact, for any $\bar{X} \leq 1.645 \cdot \sqrt{\frac{4}{100}} = 0.329$ the null will not be rejected.

Suppose however that the alternative holds. As an exercise compute β and power, $1-\beta$, for various alternatives.

Consider $\mu = .02, .08, .16, .32, .64, 1.0$

For any alternative μ $\beta(\mu) = \Pr\left(Z < \frac{.392-\mu}{\sqrt{\frac{4}{100}}}\right).$

We get $\beta(.02) = \Pr\left(Z < \frac{.392-.02}{\sqrt{\frac{4}{100}}}\right) = \Pr(Z < 1.86) =$ .9686;

$1-\beta = .0314$. So for an alternative close to the null the power is low (it is difficult to distinguish).

But $\beta(1.0) = \Pr\left(Z < \frac{.392-1.0}{\sqrt{\frac{4}{100}}}\right) = \Pr(Z < -3.04)$ is close to zero. Power close to 1.

Compute the other values and graph to get a feel for how β and the power of the test change.

In a two-sided test $\beta(\mu) = \Pr(|Z| < Z_{\alpha/2})$.

Computing β for test of a proportion.

It is more difficult to do the computation for this case because under the alternative both the mean and variance change as the proportion distribution is approximated by $N\left(p, \frac{p(1-p)}{n}\right)$ for the true p that differs depending on the (true) alternative value for p.

Example: the"grandmother hypothesis" test.

In the example probability of type II error is $\beta = \Pr($test statistic $Z < Z_a | H_1$ true).

Again, β is not constant over all possible alternatives where any of $p >$ .5 could hold. β is a function of p, $\beta(p)$.

Compute β for some specific alternatives. Set level at .01 so that $Z_\alpha = 2.33$.

This critical value implies that if $\hat{p} < 2.33\sqrt{\frac{.5\cdot.5}{126}} + .5 = 0.603\,79$, we do not reject the H_0.

For $p = .55$ we can write $\beta(.55) = \Pr(\text{test statistic } \hat{Z}_{|H_0} < Z_a | p = .55) = \Pr\left(\hat{p} < .604 | p = .55\right) = \Pr(Z < \frac{.604-.55}{\sqrt{\frac{.55\cdot.45}{126}}})$

$= \Pr(Z < 1.\,218\,4) = .88$; $1 - \beta = .12$.

For alternative $p = .6$ similarly

$\beta\,(.6) = \Pr(Z < .887) = .81$; power $=.19$;

For alternative $p = .65$ similarly

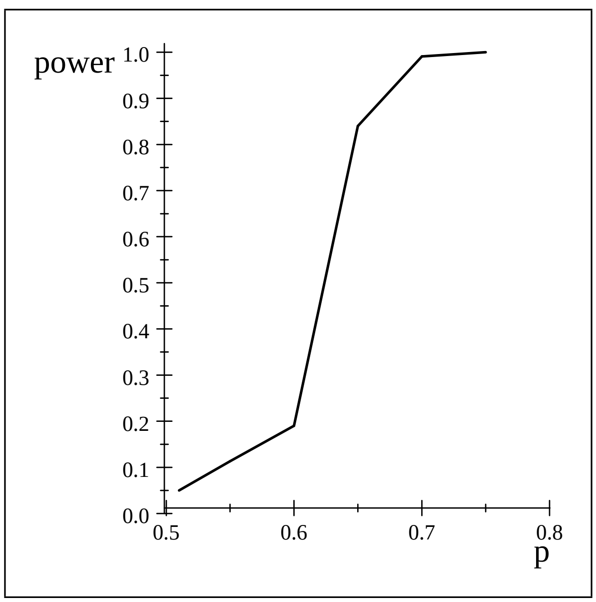
$\beta\,(.65) = \Pr(Z < -1.08) = .14$; power $=.86$;

For alternative $p = .7$ similarly

$\beta\,(.7) = \Pr(Z < -2.\,347) = .009$; power $=.991$.

So there is more power against further alternatives; power increases non-linearly from α to 1.

true p	power
.51	.05
.55	.113
.6	.19
.65	.84
.7	.991
.75	1



Tests of a difference in means

Matched pairs.

Example: Do students who rewrite the exam improve their grades on average?

In a sample of 8 who retook the exam:

row 1: make-up grade/50=X_1; row 2 MT/50=X_2; row 3 the difference: $X_1 - X_2$

30.5	30	43	24.5	28.5	35	33.5	35.5
30.0	27.5	24.5	37	32	33	25.5	31
0.5	2.5	18.5	-12.5	-3.5	2	8	4.5

$H_0 : \mu_1 - \mu_2 = 0$; $H_1 : \mu_1 - \mu_2 > 0$.

Assume first that the marks are normally distributed (unrealistic here).

Then $X_1 - X_2$ is distributed as $N(\mu_1 - \mu_2, \sigma^2_{X_1 - X_2})$.

This is **similar to testing the mean in a normal distribution**, with population variance unknown.

Estimate mean by $\bar{X}_1 - \bar{X}_2 = \frac{1}{8}20 = 2.5$; variance by $s^2 = 69.438$.

Value of statistic is $t = \frac{2.5}{\sqrt{\frac{69.438}{8}}} = 0.848\,58$; it is distributed as t_7; for $\alpha = .1$ the critical value is 1.415, the statistic is below this and we cannot reject H_0.

Sign test

There is another way to test this that does not rely on any assumption about the distribution; it is useful because here normality is not very realistic.

The sign test is a **non-parametric test of symmetry**; it tests the median rather than the mean. So now the null-hypothesis is that a student who rewrites the test is as likely to improve the mark as to see it go down.

To conduct the test record the signs of the differences: $+, +, +, -, -, +, +, +$. Under the null hypothesis what is the probability of getting 6 or more "$+$"'s out of 8 observations? Under the null-hypothesis probability of a $+$ is 0.5. and we can use the binomial to calculate the probability $\Pr(k \geq 6) = \Pr(k = 6) + \Pr(k = 7) + \Pr(k = 8) = \left(\frac{8!}{6!2!} + \frac{8!}{7!1!} + 1\right).5^8 = 0.144\,53$. This is the prob-value of the sign test. So here as well H_0 cannot b rejected at .1 level.)

Difference between means: normal distributions, known variances, independent samples.

Example. To find out whether a particular mineral in the soil improves yields it was added to the soil in one orchard and the yield of n_1 =100 apple tress there was compared to that of n_2 =150 apple trees in an orchard where the soil did not have the same level of mineral. Assume that yield of a tree in orchard 1 is X distributed normally $N(\mu_1, \sigma_1^2)$; in orchard 2 Y is distributed as $N(\mu_2, \sigma_2^2)$, with known $\sigma_1^2 = 400; \sigma_2^2 = 625$. Suppose that $\bar{X} = 110; \bar{Y} = 100$.

$H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 > 0$.

Consider difference in sample means, derive the sampling distribution:

$\bar{X} - \bar{Y}$ is normal; $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$;

$var(\bar{X} - \bar{Y}) = var\bar{X} + var\bar{Y} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

(recall that $cov(\bar{X}_1, \bar{X}_2) = 0$ since the samples are independent).

$\bar{X} - \bar{Y}$ is $N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

The test statistic then is standardized (since the variances are known it is $N(0, 1)$).

$Z = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} =$ (substituting the numbers)$= \frac{10}{\sqrt{\frac{400}{100} + \frac{625}{150}}} =$ 3.5.

To test consider either (a) given $a = .01$, or (b) compute $p - value$.

For (a) the critical value: $Z_{.01} = 2.33$. H_0 is rejected.

For (b) $p - value = \Pr(Z > 3.5) = 1 - .998 = 0.002$. So we see that we would reject even at .005 level (.5%).

Two means, normal distributions, unknown equal variances, independent samples.

For the same example change the conditions: suppose that the variances are not known but what is known that they are actually equal.

How will this differ?

We have X distributed normally $N(\mu_1, \sigma^2)$; in orchard 2 Y is distributed as $N(\mu_2, \sigma^2)$, but σ^2 is not known.

Then the sampling distribution of the difference in sample means $\bar{X} - \bar{Y}$ is still $N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$ or $N(\mu_1 - \mu_2, \frac{\sigma^2(n_1+n_2)}{n_1 n_2})$, but we need to estimate σ^2.

Different estimators are possible. One could just ignore one of the samples and estimate by s_1^2 from, say, the first sample. But that ignores the information about variance in the second sample. What is usually done is a **pooled estimator** of σ^2 that uses both samples:

$$s^2_{pooled} = \frac{1}{n_1+n_2-2}\left[\Sigma_{i=1}^{n_1}(X_i - \bar{X})^2 + \Sigma_{i=1}^{n_2}(Y_i - \bar{Y})^2\right]$$

or by using the estimated variances in each sample:

$$= \frac{1}{n_1+n_2-2}\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right].$$

Suppose all the numbers are as in the example above, but now $s_1^2 = 400$; $s_2^2 = 625$.

The standardized statistic is

$$t = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{\frac{1}{n_1+n_2-2}[(n_1-1)s_1^2 + (n_2-1)s_2^2]\frac{(n_1+n_2)}{n_1 n_2}}},$$ substituting numbers

$$\frac{10}{\sqrt{\frac{1}{248}(99 \cdot 400 + 149 \cdot 625)\frac{250}{100 \cdot 150}}} = 3.3483$$

Because we had to estimate the variance this has a $t-$distribution with $n_1 + n_2 - 2$ d.f. (-2 reflecting the fact that now two parameters - the two means - had to be estimated).

In the example the d.f. is large enough (248) that the critical values are similar to those for standard normal.

Two means, normal distributions, unknown and unequal variances, independent samples.

The sampling distribution of the difference in means is $N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Now we have no choice but to estimate σ_1^2 by s_1^2 and σ_2^2 by s_2^2 separately (cannot pool the variances).

So test statistic is

$$t = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}; \text{ numerically } = \frac{10}{\sqrt{\frac{400}{100}+\frac{625}{150}}} = 3.5.$$

The distribution of this statistic is more complicated: it is $t_{d.f.}$, but $d.f.$ follow a more elaborate formula if sample sizes differ:

$$d.f. = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2/(n_1-1)+\left(\frac{s_2^2}{n_2}\right)^2/(n_2-1)}.$$

If $n_1 = n_2$ this simplifies to $(n-1)\,\frac{(s_1^2+s_2^2)^2}{(s_1^2)^2+(s_2^2)^2}$.

So in the example: $d.f. = \dfrac{\left(\frac{400}{100}+\frac{625}{150}\right)^2}{\left(\frac{400}{100}\right)^2/99+\left(\frac{625}{150}\right)^2/149} =$ 239.79. We note that this is a fractional number (it is possible to have t with fractional d.f.); to find a critical value in the table one can take an intermediate value between the two closest integers.

Tests of variances in normal populations

Example: A dairy processing company claims that the variance of the amount of fat in the whole milk processed by the company is

no more than 0.25. You suspect this is wrong and find that a random sample of 41 milk containers has a variance of 0.27. At $\alpha = 0.05$, is there enough evidence to reject the company's claim?

Assume the population is normally distributed.

Consider X normally distributed: $N\left(\mu, \sigma^2\right)$.

The null-hypothesis, $H_0 : \sigma^2 = \sigma_0^2$.

Here,

$H_0 : \sigma^2 = .25$;

Alternatives: two-sided $H_1 : \sigma^2 \neq \sigma_0^2$, or one-sided, $H_1 : \sigma^2 > \sigma_0^2$ (or $H_1 : \sigma^2 < \sigma_0^2$).

Here this is

$H_1 : \sigma^2 \neq .25$ or a one-sided $H_1 : \sigma^2 > .25$.

Which makes more sense? Depends on the interpretation (here one-sided makes sense, that is a higher variability than claimed).

The statistic $\frac{(n-1)s^2}{\sigma_0^2}$ has a known distribution (under normality); it is χ^2_{n-1}.

Here, χ^2_{40}. The value of the statistic $\frac{40 \times .27}{.25} = 43.2$.

For $\alpha = .05$ the critical value is 55.758.

So H_0 cannot be rejected.

What can you say about the prob-value of this statistic? (It could be computed, of course, but evaluate without further computation).

P-value $>$.05.

What would be the conclusion for the two-sided test at $\alpha = .1$?

Test of equality of variances in normal populations.

Suppose that we consider two independent random samples: $\{X_1, ..., X_{n_1}\}$ from $N\left(\mu_1, \sigma_1^2\right)$ and $\{Y_1, ..., Y_{n_2}\}$ from $N(\mu_2, \sigma_2^2)$.

The null is $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$; the alternative could be one-sided or two-sided.

We know that $\frac{(n-1)s^2}{\sigma^2}$ is distributed as χ^2_{n-1}. This holds for each sample; under H_0 the denominator is the same.

Then $\frac{s_1^2}{s_2^2}$ has the distribution $\frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)}$, where the distribution in the numerator and denominator are independent (since came from independent samples). The ratio $\frac{s_1^2}{s_2^2}$ is called the $F-$ratio and the distribution is known; it is F_{n_1-1,n_2-1} - F-distribution $F_{d.f._1,d.f._2}$ with degrees of freedom $d.f._1$ called degrees of freedom of the numerator and $d.f._2$ - degrees of freedom of the denominator.

Example. We tested difference in means for two normal populations under the assumption that the variances were equal. We can now test how realistic that assumption was. In a sample of size $n_1 = 100$ we had $s_1^2 = 400$, in a sample from another distribution $n_2 = 150, s_2^2 = 625$.

$H_0 : \sigma_1^2 = \sigma_2^2$. Select a level of significance $\alpha = .05$.

Consider $\frac{s_2^2}{s_1^2}$. (This ratio because the tables of F provide upper tail values, so we use the number $>$1 for convenience).

The value of the statistic is $F = \frac{625}{400} = 1.5625$. We need the critical 5% value for 150,100 df. We can obtain it by interpolating between 1.34 (for $100, 100$ d.f.) and 1.39 (for $200, 100$ d.f.) since the table does not list those degrees of freedom. Note that this is from a table on the web

http://home.comcast.net/~sharov/PopEcol/tables/f005.html

The table in SBE lists only up to 20 d.f. in the numerator. We reject the null of equality of variances.

Difference between proportions

Example.

In 1970 a poll of university students in the US asked whether they believed that social change could be achieved by peaceful means. Out of 300 students age 18 and under 50% agreed, out of 300 students age 24 and older 69% agreed. Test whether there is no difference between the beliefs of the two groups.

$H_0 : p_1 - p_2 = 0$; $H_1 : p_1 - p_2 \neq 0$.

The sampling distribution of the difference in proportions is $\hat{p}_1 - \hat{p}_2$ is approximately normal with mean

$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$; $var(\hat{p}_1 - \hat{p}_2) = var(\hat{p}_1) + var(\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

(covariance is zero).

Distribution of $\hat{p}_1 - \hat{p}_2$ is approximately $N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$.

Then the standardized test statistic is

$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$. (this is the general form of the statistic for testing differences $p_1 - p_2$, not only that they are zero).

Under the null of equality $p_1 = p_2 = p$, so the distribution under this null is $N(0, \frac{p(1-p)}{n_1 n_2}(n_1 + n_2))$

Then the standardized test statistic is

$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1 n_2}(n_1 + n_2)}}$. The denominator can be estimated by substituting (pooled) $\hat{p}_{pooled} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$.

In the example: $\hat{p}_{pooled} = \frac{.5 \cdot 300 + .69 \cdot 300}{600} = 0.595$;

$Z = \frac{.19}{\sqrt{\frac{.595(1-.595)}{300 \cdot 300} 600}} = 4.7404$

H_0 is rejected at $\alpha = .01$.

A few problems.

1. In a sample of 400 adults and 600 teenagers 100 adults and 300 teenagers liked a TV program. Construct a 95% CI for the difference in proportions.

The sampling distribution of the difference is $N(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2})$;

The variance is estimated by $\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} = \frac{.25 \cdot .75}{400} + \frac{.5 \cdot .5}{600} = .000885$

$\sqrt{8.8542 \times 10^{-4}} =$

$\sqrt{.000885} = .029749 \approx .03$

CI is $(\hat{p}_1 - \hat{p}_2 - Z_\alpha \cdot .03, \hat{p}_1 - \hat{p}_2 + Z_\alpha \cdot .03) = (.25 - .5 - 1.96 \cdot .03, \ .25 - .5 + 1.96 \cdot .03) = (-0.3088, -0.1912)$.

Can we reject H_0 of no difference at 5% level based on this CI?

$H_0 : p_1 - p_2 = 0; H_1 : p_1 - p_2 \neq 0$

Yes.

For a one-sided alternative?

$H_1 : p_1 - p_2 > 0$

Only at .025 based on the CI.

What is the p-value here for the difference in proportions?

$\Pr(Z > \frac{.25}{.03}) \approx 0$, would reject H_0 strongly at any reasonable level.

2. (from an old exam) Suppose that in constructing a pilot ejection mechanism the manufacturer bases the

process on the assumption that the combined weight of the pilot and ejector seat is distributed normally with a mean of 130 kg and standard deviation of 25 kg. There is a view that the mean is actually lower (which if true could lead to a cost reduction in the manufacturing process).

(4)(a) State the hypotheses. Given a sample of 64 observations with average of 124 kg test the hypothesis at 5% level.

$H_0 : \mu = 130; H_1 : \mu < 130.$

Test statistic $Z = \frac{124-130}{25/8} = -1.92$. The critical value at .05 is -1.65. H_0 is rejected.

(5) (b) Find the prob-value for the sample average under the conditions in (a). Based on the prob-value test the hypothesis at 1%.

$\Pr(\bar{X} < 124) = \Pr(Z < -1.92) = .027$. Since .01<.027 the null cannot be rejected at 1%.

(5) (c) Suppose that in reality the mean is 125 kg. Find the probability of Type II error for $\alpha = .01$ and the power of the test.

So in reality $\bar{X}$ is distributed as $N(125, \frac{25^2}{64})$, so E(test statistic)=E($\frac{\bar{X}-130}{25/8}$) $= \frac{125-130}{25/8} = -1.6$, , and var(test statistic)=1 (did not change); so (test statistic+1.6) is $N(0, 1)$.

$\beta =$ Pr(Type II error)= Pr(test statistic$>-2.33|\mu = 125) =$ Pr$(Z > -2.33 + 1.6) =$ Pr$(Z > -0.73) = .767$; power is 1-$\beta = .233$.

(3) (d) Indicate how one could increase the power of the test.

Increase α. Increase sample size.

An additional question. Suppose that for safety reasons the combined weight of pilot and seat would need to be no greater than 180 with probability 99%, otherwise the mechanism design is declared unsafe.

An inspector claims that the design is unsafe and the manufacturer disputes this claim.

It is known that the variance for the pilot's weight is $\sigma_p^2 = 600$, for the weight of the seat $\sigma_s^2 = 25$; measurement were done independently for 100 pilots and 20 seats. What would the observed combined weight $\bar{X}_p + \bar{X}_s$ have to be for the manufacturer to be able to refute the inspector's claim at $\alpha = .01$?

This requires several steps.

1. First, the question is what would have to be the largest true mean weight $\mu = \mu_p + \mu_s$ that would not fail the safety constraint?

In other words for which μ

$$\Pr\left(X \geq 180 | \text{ in } N\left(\mu, \sigma_{p+s}^2\right)\right) < .01$$

holds?

Note that $\sigma^2_{p+s} = \sigma^2_p + \sigma^2_s = 625$.

This is $\Pr\left(Z > \frac{180-\mu}{25}\right) < .01$.

Then $\frac{180-\mu}{25} > 2.33$; $\mu < 180 - 25 \cdot 2.33 = 121.75$.

2. Now the inspector's claim is $H_0 : \mu \geq 121.75$ and the manufacturer wants to reject in favor of $H_1 : \mu < 121.75$ at $\alpha = .01$.

Combined weight $\bar{X} = \bar{X}_p + \bar{X}_s$ is distributed as $N\left(\mu, \frac{\sigma^2_p}{n_p} + \frac{\sigma^2_s}{n_s}\right)$, or

$$\frac{\sigma^2_p}{n_p} + \frac{\sigma^2_s}{n_s} = \frac{600}{100} + \frac{25}{20} = 7.25;$$

Then we need $\bar{X}$ for which $\Pr\left(Z < \frac{\bar{X}-121.75}{\sqrt{7.25}}\right) < .01$, or $\frac{\bar{X}-121.75}{\sqrt{7.25}} < -2.33$. Then $\bar{X} < 121.75 - 2.33 \cdot \sqrt{7.25} = 115.48$.

3. In a random sample of 100 respondents one third said that they participate in winter sports.

(5) (a) Construct a 98% two-sided confidence interval for the proportion of population that participate in winter sports.

$\hat{p} = .33$; $\hat{p}$ is approximately normally distributed with mean p and variance $\frac{p(1-p)}{n}$, then $\Pr(|\hat{p} - p| < E) = \Pr(|Z| < \frac{E}{\sqrt{\frac{p(1-p)}{n}}}) = .98 = 1 - \alpha$; $\frac{E}{\sqrt{\frac{p(1-p)}{n}}} = Z_{\alpha/2} = Z_{.01} = 2.33$ and $E = 2.33 \cdot \sqrt{\frac{.33 \cdot .67}{100}} = 0.10956$

$$
\begin{aligned}
.33 - .11 &< p < .33 + .11; \\
.22 &< p < .44.
\end{aligned}
$$

(5) (b) State what kinds of hypotheses can be tested at 2% level and which at 1% level using the result in (a)? Explain how acceptance/rejection would be decided.

At 2% any two-sided hypotheses $H_0 : p = p_0$; $H_1 : p \neq p_0$. If p_0 is inside CI in (a), H_0 is not rejected, if outside, H_0 is rejected.

At 1% one can test $H_0 : p = p_0$; against (i) $H_1 : p > p_0$ or (ii) $H_1 : p < p_0$. If p_0 is below the lower bound, H_0 is rejected in favor of (ii), if above upper, in favour of (i), otherwise, H_0 is not rejected.

(5) (c) Suppose that in another random sample of 120 40% of respondents said that they participate in sports in the summer. At 5% level test that there is no difference in the proportion of people who participate in sports in the winter and in the summer.

$H_0 : p_w - p_s = 0$; $H_1 : p_w - p_s \neq 0$.

The difference $\hat{p}_w - \hat{p}_s$ is distributed approximately as $N(p_w - p_s, \frac{p_w(1-p_w)}{n_w} + \frac{p_s(1-p_s)}{n_s})$, under H_0 this is

$N(0, \frac{p(1-p)}{n_w n_s}(n_w + n_s))$; with $p = p_w = p_s$. To estimate p use $\frac{\hat{p}_w n_w + \hat{p}_s n_s}{n_w + n_s} = \frac{.33 \cdot 100 + .4 \cdot 120}{220} = 0.368\,18$; then $\frac{p(1-p)}{n_w n_s}(n_w + n_s)$ is estimated by $\frac{.37 \cdot .63}{100 \cdot 120}(220) = .0043$; test statistic is $Z = \frac{-.067}{\sqrt{.0043}} = -1.\,021\,7$. the critical value is -1.96 and the null cannot be rejected.

4. A new company offers insurance against theft of home computers. It is known that every year 1.5% of computer users have their computers stolen. The mean insured value of a home computer is 1000$ and the company decides to set the premiums at 15.50$/year.

(5) (a) In the first year 1.5% of the 1000 clients report a stolen computer. However, the average claim was higher than expected and as a result the company made a loss. What may be suspected of the distribution of value of the stolen computers versus the overall distribution of values of computers? Assuming that the distribution of values of stolen computers is normal and that the average claim was \$1050 with estimated standard deviation of \$100 test

the hypothesis that the mean value of a stolen computer is \$1000; test at 5% and 1% levels.

The value of a stolen computer is higher on average.

$H_0 : \mu_s = 1000$; $H_1 : \mu_s > 1000$.

$t_{14} = \frac{50}{\sqrt{10000/15}} = 1.9365$; H_0 is rejected at 5% level since the critical value is 1.761, but not at 1% for 2.624.

(3) (b) How would your answer in (a) be affected if normality could not be assumed?

We would not know the distribution of the test statistic; the sample size of 15 is not large and the t approximation may be poor; the results will not be reliable.